

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:05:01 ; Search time 519.52 Seconds  
(without alignments)  
531.373 Million cell updates/sec

Title: US-09-867-034-2

Perfect score: 322

Sequence: 1 gcaacgacacaaaccttga.....cttcaagtgtcgaagact 322

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No:	Score	Query Match	Length	DB ID	Description
1	321	99.7	525	19	AAV29048
2	321	99.7	866	19	AAV29047
3	321	99.7	875	19	AAV19155
4	321	99.7	875	22	AAV63314
5	321	99.7	881	19	AAV59320
6	321	99.7	920	21	AAV58380
7	321	99.7	943	22	AAV35026
8	321	99.7	1018	20	AAZ77486
9	321	99.7	1020	20	AAZ52865
10	321	99.7	1689	20	AAZ40803
11	310	96.3	529	22	AAH55526

12	309	96.0	778	22	AAFA4884	Human breast cancer
13	302.2	93.9	940	20	AAZ40846	Secreted protein E
14	275	85.4	386	19	AAV19188	Human XAG growth f
15	275	85.4	386	22	AAV63347	HuXAG-1 related ES
16	263.4	81.8	401	21	AAV7753	cDNA encoding huma
17	263.4	81.8	401	22	AAI28491	Colon tumour relat
18	249	77.3	409	21	AAV7903	cDNA encoding huma
19	249	77.3	409	22	AAI28641	Colon tumour relat
20	246.8	76.6	793	21	AAZ24578	Human lung tumor a
21	246.8	76.6	793	20	AAV59329	Human lung cancer-
22	233	72.4	525	19	AAV59329	Degenerate DNA seq
23	205.8	63.9	506	19	AAV19159	Human XAG growth f
24	205.8	63.9	506	22	AAV63334	HuXAG-1 related cd
25	188	58.4	468	21	AAV00115	Human secreted pro
26	167.8	52.1	321	22	AAV17730	Human breast cancer
27	159.4	49.5	235	22	AAV68475	Human lung tumour
28	159.4	49.5	301	21	AAV7841	cDNA encoding huma
29	159.4	49.5	301	22	AAI28579	Colon tumour relat
30	157.2	48.8	501	22	AAV54130	Breast cancer prot
31	157.2	48.8	572	20	AAV88876	EST clone FTY404.
32	157.2	48.8	640	22	AAI29081	Colon tumour relat
33	157.2	48.8	728	22	AAV86999	Nucleotide sequenc
34	157.2	48.8	732	19	AAV19157	Human XAG growth f
35	157.2	48.8	732	22	AAV63316	Human huXAG-3/CCSG
36	157.2	48.8	749	21	AAV54129	Breast cancer prot
37	153	47.5	404	21	AAV65978	Human lung cancer-
38	151	46.9	398	20	AAV41145	Human secreted pro
39	149.4	46.4	420	22	AAV98692	Human ovarian can
40	148	45.0	520	21	AAZ42286	Human 5' EST isol
41	146.2	45.4	896	22	AAH33168	Human colon cancer
42	144.4	44.8	171	22	AAV1924	Human breast cancer
43	136	42.2	462	21	AAV7985	cDNA encoding huma
44	136	42.2	462	22	AAI28723	Colon tumour relat
45	95.8	29.8	401	22	AAI29090	Colon tumour relat

#### ALIGNMENTS

RESULT 1	AAV29048	standard; cDNA; 525 BP.
ID	AAV29048	
XX	AAV29048	
AC	AAV29048	
XX	21-AUG-1998	(first entry)
XX	Open reading frame	human protein comprising secretory signal 9.
DE	Human protein; secretory signal; nutritional source; cytokine;	
XX	immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;	
KW	chemokine; thrombolytic; anti-inflammatory; inhibition;	
KW	stomach cancer cell; ds.	
XX		
XX	Homo sapiens.	
OS		
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..525
FT		/*tag= a
FT		/product= "human protein comprising secretory
FT		signal"
PN	W09811217-A2.	
XX		
PD	19-MAR-1998.	
XX		
PF	12-SEP-1997;	97WO-JP03239.
XX		
PR	13-SEP-1996;	96JP-0243060.
XX		
PA	(PROT-) PROTEGENE INC.	
XX	(SAGA) SAGAMI CHEM RES CENTRE.	

PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 XX  
 DR MPI; 1998-207380/18.  
 DR P-PSDB; AAM37872.  
 XX  
 PT Human proteins with secretory signal sequences - used to treat  
 PT immune deficiencies, infections, tumours, and haematopoietic  
 PT disorders, etc.  
 XX  
 PS Claim 3; Pages 88; 131pp; English.  
 XX  
 CC This is the nucleotide sequence of the open reading frame of a novel  
 CC human protein comprising a secretory signal (AAV29047), isolated from  
 CC stomach cancer cells. Its proteins can be used as nutritional sources  
 CC or supplements. The proteins may also have cytokine functions,  
 CC immune modulating functions, haematopoiesis regulating actions,  
 CC activin/inhibin regulating activity, chemotactic/chemokine activity,  
 CC haemostatic and thrombolytic activity, receptor/ligand activity,  
 CC anti-inflammatory activity, tumour inhibition activity.  
 CC  
 SQ Sequence 525 BP; 162 A; 131 C; 103 G; 129 T; 0 other;

Query Match 99.7%; Score 321; DB 19; Length 525;  
 Best Local Similarity 99.7%; Pred. No. 2.3e-86;  
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 60  
 |||||||  
 DB 198 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 257  
 |||||||  
 OY 61 tttaagaaagtgttgcgtgaaataaagaatccagaatctggcagagcagttgtcct 120  
 |||||||  
 DB 258 tttaagaaagtgttgcgtgaaataaagaatccagaatctggcagagcagttgtcct 317  
 |||||||  
 OY 121 cctcaatctggtttatgaaacactgacacaccccttcctcctgagtgccagtatgtccc 180  
 |||||||  
 DB 318 cctcaatctggtttatgaaacactgacacaccccttcctcctgagtgccagtatgtccc 377  
 |||||||  
 OY 181 cagattatgttctgtagccatctcgcagctgagcagatccatcctggaatattc 240  
 |||||||  
 DB 378 cagattatgttctgtagccatctcgcagctgagcagatccatcctggaatattc 437  
 |||||||  
 OY 241 aaancgtctatgctacgacacgtgcagatagcgtctgttgccttgacacatgaa 300  
 |||||||  
 DB 438 aaacgctctatgctacgacacgtgcagatagcgtctgttgccttgacacatgaa 497  
 |||||||  
 OY 301 agctctcaagttgctgaagact 322  
 |||||||  
 DB 498 agctctcaagttgctgaagact 519  
 |||||||

RESULT 2  
 AAV29047  
 ID AAV29047 standard; cDNA; 866 BP.  
 XX  
 AC AAV29047;  
 XX  
 DT 21-AUG-1998 (first entry)  
 XX  
 DE Human protein comprising secretory signal nucleotide sequence 9.  
 XX  
 KW Human protein; secretory signal; nutritional source; cytokine;  
 KW immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;  
 KW chemokine; thrombolytic; anti-inflammatory; inhibition;  
 KW stomach cancer cell; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 73..600  
 FT /tag= a  
 FT /product= "human protein comprising secretory

FT  
 XX  
 XX  
 PN MO9811217-A2.  
 XX  
 PD 19-MAR-1998.  
 XX  
 PE 12-SEP-1997; 97MO-JP03239.  
 XX  
 PR 13-SEP-1996; 96JP-0243060.  
 XX  
 PA (PROT-) PROTEGENE INC.  
 PA (SAGA) SAGAMI CHEM RES CENTRE.  
 XX  
 PI Kato S, Kobayashi M, Sekine S, Yamaguchi T;  
 DR MPI; 1998-207380/18.  
 DR P-PSDB; AAM37872.  
 XX  
 PT Human proteins with secretory signal sequences - used to treat  
 PT immune deficiencies, infections, tumours, and haematopoietic  
 PT disorders, etc.  
 XX  
 PS Claim 4; Pages 114-116; 131pp; English.  
 XX  
 CC This is the nucleotide sequence of a novel human protein comprising  
 CC a secretory signal isolated from stomach cancer cells. Its proteins  
 CC can be used as nutritional sources or supplements. The proteins may  
 CC also have cytokine functions, immune modulating functions,  
 CC haematopoiesis regulating activity, activin/inhibin regulating  
 CC activity, chemotactic/chemokine activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity.  
 CC  
 SQ Sequence 866 BP; 275 A; 189 C; 178 G; 224 T; 0 other;

Query Match 99.7%; Score 321; DB 19; Length 866;  
 Best Local Similarity 99.7%; Pred. No. 2.7e-86;  
 Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 60  
 |||||||  
 DB 270 gacaagcaaaaccccttgatgattatcactcacttgatgagtgccacacagtcgaagc 329  
 |||||||  
 OY 61 tttaagaaagtgttgcgtgaaataaagaatccagaatctggcagagcagttgtcct 120  
 |||||||  
 DB 330 tttaagaaagtgttgcgtgaaataaagaatccagaatctggcagagcagttgtcct 389  
 |||||||  
 OY 121 cctcaatctggtttatgaaacactgacacaccccttcctcctgagtgccagtatgtccc 180  
 |||||||  
 DB 390 cctcaatctggtttatgaaacactgacacaccccttcctcctgagtgccagtatgtccc 449  
 |||||||  
 OY 181 cagattatgttctgtagccatctcgcagctgagcagatccatcctggaatattc 240  
 |||||||  
 DB 450 cagattatgttctgtagccatctcgcagctgagcagatccatcctggaatattc 509  
 |||||||  
 OY 241 aaancgtctatgctacgacacgtgcagatagcgtctgttgccttgacacatgaa 300  
 |||||||  
 DB 510 aaacgctctatgctacgacacgtgcagatagcgtctgttgccttgacacatgaa 569  
 |||||||  
 OY 301 agctctcaagttgctgaagact 322  
 |||||||  
 DB 570 agctctcaagttgctgaagact 591  
 |||||||

RESULT 3  
 AAV19155  
 ID AAV19155 standard; cDNA; 875 BP.  
 XX  
 AC AAV19155;  
 XX  
 DT 28-AUG-1998 (first entry)  
 XX

DE	Human XAG growth factor huXAG-1 cDNA.
KW	huXAG-1; XAG; growth factor; colon cancer-specific gene;
KW	tumour marker; breast disease; liver disease; lung disease;
KW	emphysema; wound healing; diagnosis; therapy; human; ds.
XX	
OS	Homo sapiens.
XX	
XX	
FH	Key
FT	Location/Qualifiers
FT	CDS
FT	71..598
FT	/*tag= a
FT	sig_peptide
FT	71..130
FT	/*tag= b
FT	mat_peptide
FT	131..595
FT	/*tag= c
XX	
PN	WO9807749-A1.
XX	
XX	
PD	26-FEB-1998.
XX	
XX	
PE	22-AUG-1997; 97WO-US14139.
XX	
PR	23-AUG-1996; 96WO-US13766.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Dillon PJ, Ebner R, Endress GA, Yu G;
DR	WPI; 1998-169093/15.
DR	P-PSDB; AAM37844.
XX	
PT	New isolated human XAG growth factor(s) - used to develop products
PT	for treating e.g. liver, lung or breast diseases or
XX	hyperproliferative disorders, e.g. cancer.
PS	Claim 5; Fig 1; 141pp; English.
XX	
XX	This cDNA clone codes for huXAG-1 (see AAM37844), a member of a novel
CC	family of human growth factors also including huXAG-2 (see AAM37845)
CC	and huXAG-3 (see AAM37846). These proteins share homology with the
CC	XAG protein of Xenopus laevis, which is involved in embryogenesis
CC	and is expressed in adult tissue. Expression of huXAG-1 has been
CC	discovered in colon cancer tissue, with no corresponding expression
CC	in healthy colon tissue. The huXAG-1 gene, also designated
CC	colon cancer-specific gene (CCSG), provides a molecular marker for
CC	colon cancer. huXAG-1 cDNA was isolated from a cDNA library
CC	derived from human colon cancer tissue. Vectors, host cells,
CC	antibodies, and screening methods for identifying agonists and
CC	antagonists of huXAG-1 are provided. HuXAG polypeptides are growth
CC	factors and can be used to stimulate proliferation of cells. They
CC	can be used to stimulate the proliferation and differentiation of
CC	hepatocytes to alleviate or treat liver diseases and pathologies
CC	such as fulminant liver failure caused by cirrhosis, liver damage
CC	caused by viral hepatitis and toxic substances. They can also be
CC	used to stimulate or promote liver regeneration, e.g. after
CC	surgery. They can also be used to prevent and heal damage to the
CC	lungs caused by various pathological states. They can be used to
CC	stimulate proliferation and differentiation and promote the repair
CC	of alveoli and bronchiolar epithelium to prevent, attenuate, or
CC	treat acute or chronic lung damage, e.g. emphysema, which results
CC	in the progressive loss of alveoli, and inhalation injuries, e.g.
CC	resulting from smoke inhalation and burns, that cause necrosis of
CC	the bronchiolar epithelium and alveoli. They can also be used to
CC	stimulate the proliferation and differentiation of breast tissue
CC	and could therefore be used to promote healing of breast tissue
CC	injury due to surgery, trauma or cancer. Antagonists can be used
CC	to treat hyperproliferative disorders, including cancer, in
CC	particular hepatocellular carcinoma, osteosarcoma, breast cancer,
CC	or colon cancer. The products can also be used for detection and
CC	diagnosis.
XX	
Sequence	875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;

Query Match	99.7%	Score 321:	DB 19:	Length 875:
Best Local Similarity	99.7%	Pred. No. 2.7e-86:		
Matches 321:	Conservative 0:	Mismatches 1:	Indels 0:	Gaps 0:
QY 1	gacaaagcaacaacccctgatactgatatcatcaacttggatgagtcgcccacagccaagc	60		
Db 268	gacaaagcaacaacccctgatactgatatcatcaacttggatgagtcgcccacagccaagc	327		
QY 61	tttaagaagaatgcttctgtaaaataaagaatccagaatctggcagagcagttgtccct	120		
Db 328	tttaagaagaatgcttctgtaaaataaagaatccagaatctggcagagcagttgtccct	387		
QY 121	cctaactgtgtttatagtaaacactgcgaacaaacactttctcctgatagcagatgtccc	180		
Db 388	cctaactgtgtttatagtaaacactgcgaacaaacactttctcctgatagcagatgtccc	447		
QY 181	cagaatattgtttgttgaccatcctctacagtttagacgcatatcacatgtgaagaatc	240		
Db 448	cagaatattgtttgttgaccatcctctacagtttagacgcatatcacatgtgaagaatc	507		
QY 241	aaanagctctctatgcttaagcaacctgcagatacagctctgttgccttgaacaatgaaga	300		
Db 508	aaatgcctctcatgcttaagcaacctgcagatacagctctgttgccttgaacaatgaaga	567		
QY 301	agctctcaagttgtctgaagact	322		
Db 568	agctctcaagttgtctgaagact	569		

	RESULT	4	
XX	AA63314		
XX	AA63314 standard; cDNA;	875 BP.	
AC	AA63314;		
XX			
DT	04-MAY-2001	(first entry)	
XX			
DE	Human huxAG-1/CCSG colon cancer specific gene cDNA.		
XX			
RW	Human; growth factor; huxAG-1; colon cancer specific gene; CCSG;		
KW	cell proliferation; liver disease; fulminant liver failure; cirrhosis;		
KW	hepatitis; cancer; colon cancer; colorectal carcinoma; ss.		
XX			
OS	Homo sapiens.		
XX			
FN	US6171816-B1.		
XX			
PD	09-JAN-2001.		
XX			
PF	22-AUG-1997;	97US-0916576.	
XX			
FR	23-AUG-1996;	96US-0024347.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
XX			
PI	Yu G, Dillon PJ, Ebner R, Endress GA;		
DR	WPL: 2001-136872/14.		
DR	P-PADB: AAB72203.		
XX			
PT	Novel human growth factor polypeptide useful for diagnosing and		
PT	treating colon cancer and liver diseases, to prevent and heal damage to		
PT	the lungs and for identifying modulators of therapeutic use -		
XX			
PS	Claim 45; Fig 1; 59pp; English.		
XX			
CC	This invention relates to a human growth factor polypeptide huxag-1 also		
CC	known as a colon cancer specific gene (CCSG). HUXAG-1 stimulates cell		
CC	proliferation as a growth factor. The HUXAG-1 protein is useful for		
CC	identifying compounds capable of enhancing or inhibiting cellular		
CC	response induced by huxAG-1. The protein is also useful for stimulating		
CC	proliferation of cells e.g. colon, breast, liver and lung cells, and		

CC hepatocytes. It is useful for alleviating or treating liver diseases and  
CC pathologies such as fulminant liver failure caused by cirrhosis, liver  
CC damage caused by viral hepatitis and toxic substances, for preventing and  
CC treating damage to the lungs caused by various pathological strains and  
CC for promoting healing of breast tissue injury due to surgery, trauma or  
CC cancer. huXAG-1 and the identified antagonist are useful for treating  
CC cancer, in particular colon cancer. Detecting altered levels of huXAG-1  
CC and its polynucleotides are useful for diagnosing or detecting cancer in  
CC mammals. The gene encoding huXAG-1 is useful for monitoring human  
CC colorectal carcinoma. huXAG-1 nucleic acid molecules are also useful for  
CC chromosome identification. The present sequence represents cDNA encoding  
CC huXAG-1.  
XX  
SQ

Sequence 875 BP; 283 A; 189 C; 180 G; 223 T; 0 other;

Query Match 99.7%; Score 321; DB 22; Length 875;  
Best Local Similarity 99.7%; Pred. No. 2.7e-86;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaacacccttgatgattatcacttgatgagtgccacacagtcacg 60  
DB 268 gacaagcaaacacccttgatgattatcacttgatgagtgccacacagtcacg 327  
QY 61 tttaagaaagtgcttgctgaataaagaatccagaatctggcagagcagttgtcct 120  
DB 328 tttaagaaagtgcttgctgaataaagaatccagaatctggcagagcagttgtcct 387  
QY 121 cctcaatcgtgttatgaacacactgacaacaccccttcctcgaatggcagatgtgcc 180  
DB 388 cctcaatcgtgttatgaacacactgacaacaccccttcctcgaatggcagatgtgcc 447  
QY 181 caggaattatgttctgacaccatctgacagttagagccgatacactggaagatattc 240  
DB 448 caggaattatgttctgacaccatctgacagttagagccgatacactggaagatattc 507  
QY 241 aaangctctctatgcttaagaacctgcagatacagctctgttcttgacacaatgaaga 300  
DB 508 aaatcgtctctatgcttaagaacctgcagatacagctctgttcttgacacaatgaaga 567  
QY 301 agctccaagtgtcgtgaagact 322  
DB 568 agctccaagtgtcgtgaagact 589

RESULT 5  
AAV59320 standard; cDNA; 881 BP.

XX AAV59320;

XX 21-DEC-1998 (first entry)

XX Nucleotide sequence encoding zsig10 polypeptide.

XX ss; human; mucous-mediated function; adhesion; tumour metastasis;  
KW bacterial colonisation; microbial infection; AIDS; cystic fibrosis;  
KW chronic obstructive pulmonary disease; asthma; Crohn's disease;  
KW sinonasal inflammatory disease; inflammatory bowel disease; bronchitis.  
XX

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 63..590

XX FT /\*tag= a

XX FT /product= "zsig10"

XX W09841627-A1.

XX 24-SEP-1998.

XX 18-MAR-1998; 98WO-US05251.

PR 19-MAR-1997; 9705-0039631.

XX (ZYMO ) ZYMOGENETICS INC.

XX Shepard PO;

XX WPI; 1998-531566/45.

XX P-PSDB; AAW77365.

XX New isolated mucous-associated polypeptide, zsig10 - used to develop  
PT products for treating e.g. tumour metastasis, microbial infections,  
PT cystic fibrosis, asthma, bronchitis or inflammatory bowel disease  
XX

XX Claim 37; Page 80-81; 109pp; English.

CC The human polypeptide zsig10 is involved in mucous-mediated functions  
CC such as adhesion. The products of the invention can be used in the study  
CC and treatment of e.g. tumour metastasis, bacterial colonisation,  
CC susceptibility to and persistence of infection, microbial infections,  
CC AIDS, cystic fibrosis, chronic obstructive pulmonary disease, asthma,  
CC sinonasal inflammatory disease, inflammatory bowel disease, bronchitis,  
CC or Crohn's disease. The products can also be used for detection,  
CC diagnosis and drug screening.

SQ Sequence 881 BP; 293 A; 185 C; 180 G; 223 T; 0 other;

Query Match 99.7%; Score 321; DB 19; Length 881;  
Best Local Similarity 99.7%; Pred. No. 2.8e-86;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaacacccttgatgattatcacttgatgagtgccacacagtcacg 60  
DB 260 gacaagcaaacacccttgatgattatcacttgatgagtgccacacagtcacg 319  
QY 61 tttaagaaagtgcttgctgaataaagaatccagaatctggcagagcagttgtcct 120  
DB 320 tttaagaaagtgcttgctgaataaagaatccagaatctggcagagcagttgtcct 379  
QY 121 cctcaatcgtgttatgaacacactgacaacaccccttcctcgaatggcagatgtgcc 180  
DB 380 cctcaatcgtgttatgaacacactgacaacaccccttcctcgaatggcagatgtgcc 439  
QY 181 caggaattatgttctgacaccatctgacagttagagccgatacactggaagatattc 240  
DB 440 caggaattatgttctgacaccatctgacagttagagccgatacactggaagatattc 499  
QY 241 aaangctctctatgcttaagaacctgcagatacagctctgttcttgacacaatgaaga 300  
DB 500 aaatcgtctctatgcttaagaacctgcagatacagctctgttcttgacacaatgaaga 559  
QY 301 agctccaagtgtcgtgaagact 322  
DB 560 agctccaagtgtcgtgaagact 581

RESULT 6  
AAC58380 standard; cDNA; 920 BP.

XX AAC58380;

XX 29-JAN-2001 (first entry)

XX Human PRO1030 nucleotide sequence SEQ ID NO:39.

XX Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth;  
KW proliferation; tumorigenesis; identification; cancer; cytostatic;  
KW neutrotropic; neuroprotective; antiinflammatory; immunosuppressive;  
KW immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy;  
KW neuronal disorder; glial disorder; astrocytal disorder; angiogenic;  
KW hypothalamic disorder; glandular disorder; macrophagal disorder;  
KW epithelial disorder; stromal disorder; blastocellic disorder;



KW Inflammatory disorder; immunologic disorder; ss.  
XX Homo sapiens.  
OS  
XX WO200053755-A2..  
XX  
PD 14-SEP-2000.  
XX  
XX 06-JAN-2000; 2000WO-US00376.  
XX  
XX 08-MAR-1999; 99MO-US05028.  
PR 02-JUN-1999; 99MO-US12252.  
PR 23-JUN-1999; 99US-0141037.  
PR 07-JUL-1999; 99US-0143048.  
PR 26-JUL-1999; 99US-0145698.  
PR 30-NOV-1999; 99MO-US28313.  
PR 20-DEC-1999; 99MO-US30911.  
PR 05-JAN-2000; 2000WO-US00219.  
XX  
XX (GETH ) GENENTECH INC.  
PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA;  
PI Watanabe CK, Wood WI;  
DR WPI: 2000-572270/53.  
DR P-PSDB; AAB24070.  
XX  
PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the  
PT treatment, diagnosis and prevention of cancer -  
XX  
PS Claim 50; Fig 27; 286pp; English.  
XX  
XX The present invention describes an isolated antibody that binds to  
CC one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO355,  
CC PRO619, PRO717, PRO805, PRO830, PRO848, PRO943, PRO1005, PRO1009,  
CC PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184,  
CC PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094,  
CC PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell  
CC growth. The PRO polypeptides and nucleotides are useful in the  
CC treatment, diagnosis and prevention of cancer. The antibodies and other  
CC anti-tumour compounds may be used to treat various conditions, including  
CC those characterised by overexpression and/or activation of the amplified  
CC PRO genes. Exemplary conditions or disorders to be treated with such  
CC antibodies and other compounds include benign or malignant tumours  
CC (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian,  
CC colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic  
CC carcinomas, sarcomas, glioblastomas, and various head and neck tumours),  
CC leukemias and lymphoid malignancies, other disorders such as neuronal,  
CC glioma, astrocytoma, hypopharyngeal and other glandular, macrophagal,  
CC epithelial, stromal and blastocoeleic disorders, and inflammatory,  
CC angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR  
CC primers and hybridisation probes used in the isolation of the human PRO  
CC sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human  
CC PRO polynucleotide and protein sequences given in the exemplification of  
XX the present invention.  
XX  
XX Sequence 920 BP; 296 A; 198 C; 195 G; 231 T; 0 other:  
SQ

Query Match 99.7%; Score 321, DB 21; Length 920;  
Best Local Similarity 99.7%; Pred. No. 2.8e-86;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
XX  
QY 1 gaaagaaacacccctgattcattcattcattgtagagcgcacacacagc 60  
DB 345 gacagagaaacacccctgattcattcattcattgtagagcgcacacacagc 404  
QY 61 tttaagaaagtgttctgtaaaataaagaatccagaatgtagcagcagttgtcct 120  
DB 405 tttaagaaagtgttctgtaaaataaagaatccagaatgtagcagcagttgtcct 464  
QY 121 cctcaatctgttattatgaacacagcacaacaccttctctctgtagccagatgtctcc 180  
XX

DB 465 cctcaatctgttattatgaacacagcacaacaccttctctctgtagccagatgtctcc 524  
QY 181 cagattatgttctgtgacccatctctgacagtttagagccgatatcactggaatattc 240  
DB 525 cagattatgttctgtgacccatctctgacagtttagagccgatatcactggaatattc 584  
QY 241 aaancgtctcatgcttgaacacctgagatcacagctctgttctgtgacacataaagaa 300  
DB 585 aaatcgtctcatgcttgaacacctgagatcacagctctgttctgtgacacataaagaa 644  
QY 301 agcttcaagtgtcgtgaagact 322  
DB 645 agcttcaagtgtcgtgaagact 666  
XX

RESULT 7  
AAH35026 standard; CDNA: 943 BP.  
XX  
XX AAH35026;  
AC  
XX 03-SEP-2001 (first entry)  
DT  
XX  
XX Human colon cancer antigen encoding CDNA SEQ ID NO:2108.  
DE  
XX Human: colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
XX  
XX Homo sapiens.  
OS  
XX WO200122920-A2.  
XX  
XX 05-APR-2001.  
PD  
XX 28-SEP-2000; 2000WO-US26524.  
PF  
XX 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
FA  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
PI WPI: 2001-235357/24.  
PT P-PSDB: AAG75621.  
DR  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
PT useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
XX Claim 1; Page 3593; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG75614 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated P,  
CC by inserting the nucleic acids into a host cell and culturing the cell  
CC to express the proteins. N and P can be used in the prevention, diagnosis  
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
CC present invention.  
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
CC missing at time of publication, meaning no sequences are present for  
CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
XX Sequence 943 BP; 311 A; 199 C; 199 G; 229 T; 5 other;  
SQ

Query Match 99.7%; Score 321; DB 22; Length 943;  
Best Local Similarity 99.7%; Pred. No. 2.8e-86;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gaaagcaaaaccccttgatgattatcatcacttgatgagtgccacacagtcagc 60  
|||||  
DB 280 gaaagcaaaaccccttgatgattatcatcacttgatgagtgccacacagtcagc 339  
61 ttcaagaaagtgcttgcctgaataaagaatccagaatitggcagagagttgtcct 120  
DB 340 tttaagaaagtgcttgcctgaataaagaatccagaatitggcagagagttgtcct 399  
QY 121 cctcaatctggtttatgaacaactgacaacaccccttctccatgagcagatgtccc 180  
|||||  
DB 400 cctcaatctggtttatgaacaactgacaacaccccttctccatgagcagatgtccc 459  
QY 181 cagatattgcttgcctgaacacacacacacacacacacacacacacacacacac 240  
|||||  
DB 460 cagatattgcttgcctgaacacacacacacacacacacacacacacacacacac 519  
QY 241 aaagctctctatgcttaagcaacctgcagatagagctgtgcttgacacacatgaaga 300  
|||  
DB 520 aaagctctctatgcttaagcaacctgcagatagagctgtgcttgacacacatgaaga 579  
QY 301 agctctcaagttgctgaagact 322  
|||||  
DB 580 agctctcaagttgctgaagact 601

## RESULT 8

AAZ77486  
ID AAZ77486 standard; cDNA; 1018 BP.

AC AAZ77486;

DT 10-APR-2000 (first entry)

DE Human ovarian tumor cDNA library derived EST fragment 37.

XX Expressed sequence tag; EST; human; ovarian tumor; anticancer;  
KW gene therapy; treatment; ss.

OS Homo sapiens.

PN DE19817557-A1.

PD 21-OCT-1999.

PF 09-APR-1998; 98DE-1017557.

PR 09-APR-1998; 98DE-1017557.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR MPI; 1999-591920/51.

XX P-PSDB; AAY76590.

PT New nucleic acid sequences expressed in ovarian, and some other, cancer  
PT tissues, and derived polypeptides, for treatment of ovarian cancer and  
PT identification of therapeutic agents

PS Claim 3; Page 167; 310pp; German.

XX This invention describes novel nucleic acid (cDNA) sequences (A) which  
CC have anticancer activity and are highly expressed in ovarian tumor  
CC tissue (and some also in testis and breast cancer tissue). The products  
CC of the invention can be used for gene therapy. (A) are used (i) for  
CC recombinant expression of polypeptides (B) and (ii) to isolate complete  
CC genes. (B) are used (i) to identify agents suitable for treatment of  
CC ovarian cancer; (ii) directly for treating this form of cancer

CC (including expression from gene therapy vectors) and (iii) for generation  
CC of specific antibodies. (A) are identified by assembling ESTs (expressed  
CC sequence tags) from a particular tissue type before comparison of  
CC expression patterns. This allows a significantly longer fragment of the  
CC gene to be revealed, so should reduce the number of failures associated  
CC with the fact that ESTs from different libraries may represent different  
CC parts of the same unknown gene, distorting the estimated frequency of  
CC occurrence in a particular tissue. AAZ77450-777572 represent the human  
CC ovarian tumor cDNA library derived EST fragments described in the method  
CC of the invention and encode the protein fragments represented in  
CC AAY76505-Y76638.

SQ Sequence 1018 BP; 343 A; 212 C; 216 G; 247 T; 0 other;

Query Match 99.7%; Score 321; DB 20; Length 1018;  
Best Local Similarity 99.7%; Pred. No. 2.9e-86;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gaaagcaaaaccccttgatgattatcatcacttgatgagtgccacacagtcagc 60  
|||||  
DB 345 gaaagcaaaaccccttgatgattatcatcacttgatgagtgccacacagtcagc 404  
61 tttaagaaagtgcttgcctgaataaagaatccagaatitggcagagagttgtcct 120  
|||||  
DB 405 tttaagaaagtgcttgcctgaataaagaatccagaatitggcagagagttgtcct 464  
QY 121 cctcaatctggtttatgaacaactgacaacaccccttctccatgagcagatgtccc 180  
|||||  
DB 465 cctcaatctggtttatgaacaactgacaacaccccttctccatgagcagatgtccc 524  
QY 181 cagatattgcttgcctgaacacacacacacacacacacacacacacacacacac 240  
|||||  
DB 525 cagatattgcttgcctgaacacacacacacacacacacacacacacacacacac 584  
QY 241 aaagctctctatgcttaagcaacctgcagatagagctgtgcttgacacacatgaaga 300  
|||  
DB 585 aaagctctctatgcttaagcaacctgcagatagagctgtgcttgacacacatgaaga 644  
QY 301 agctctcaagttgctgaagact 322  
|||||  
DB 645 agctctcaagttgctgaagact 666

## RESULT 9

AAZ52865  
ID AAZ52865 standard; cDNA; 1020 BP.

AC AAZ52865;

DT 14-MAR-2000 (first entry)

DE Human prostate tumor cDNA library derived EST fragment #8.

XX Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

KW treatment; ds.

OS Homo sapiens.

PN DE19820190-A1.

PD 04-NOV-1999.

PF 28-APR-1998; 98DE-1020190.

PR 28-APR-1998; 98DE-1020190.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;

DR MPI; 1999-621386/54.

XX P-PSDB; AAY73835, AAY73836, AAY73837.

```

XX New human nucleic acid sequences from pancreatic tumors, and related
PT proteins
XX
XX Claim 2; Page 188; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acids derived from human pancreatic tumor tissue which have cytostatic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AA252858-253014
CC represent expressed sequence tag (EST) fragments derived from a human
CC pancreatic tumor cDNA library and which encode the proteins represented
CC in AA252858-253014.
XX
XX Sequence 1020 BP; 341 A; 214 C; 218 G; 247 T; 0 other;
XX
Query Match          99.7%; Score 321; DB 20; Length 1020;
Best Local Similarity 99.7%; Pred. No. 2.9e-86;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacaagacaacacctgattatcattcatttgatgagcgccacacagtcacg 60
   |||||||
Db 347 gacaagacaacacctgattatcattcatttgatgagcgccacacagtcacg 406
QY 61 tttaagaagaagtgttgcgtaaaataagaataccagaataatggcagagcattgtcct 120
   |||||||
Db 407 tttaagaagaagtgttgcgtaaaataagaataccagaataatggcagagcattgtcct 466
QY 121 cctcaatctggtttatgaacaactgacacacaccttctctcgtatggcagatgtctcc 180
   |||||||
Db 467 cctcaatctggtttatgaacaactgacacacaccttctctcgtatggcagatgtctcc 526
QY 181 cagattatgtttgttaccatctcgcagctgttagagccgatatcattgaagaatattc 240
   |||||||
Db 527 cagattatgtttgttaccatctcgcagctgttagagccgatatcattgaagaatattc 586
QY 241 aaanagctctatgcttaccagacctgcagatcacagctcgtgtgttgcgacaacatgaaga 300
   |||||||
Db 587 aaacgctctatgcttaccagacctgcagatcacagctcgtgtgttgcgacaacatgaaga 646
QY 301 agctctcaagtgtcgtgaagact 322
   |||||||
Db 647 agctctcaagtgtcgtgaagact 668
XX
RESULT 10
ID AA240803 standard; DNA; 1689 BP.
XX
XX AA240803;
XX
DT 18-JAN-2000 (first entry)
XX
DE Secreted protein EST coding sequence 108-008-5-0-A6-FL.
XX
XX Secreted protein; fingerprint identification technique;
XX chromosome mapping; human; hereditary disease; diagnosis; cancer;
XX hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
XX autoimmune disease; rheumatic disease; embryonic disorder; myopathy;
XX renal injury; amino aciduria; hypoglycaemia; male rat infertility;
XX hypertension; ss.
XX
XX Homo sapiens.
XX
XX MO9940189-A2.
XX
XX 12-AUG-1999.
XX
XX 09-FEB-1999; 99WC-IB00282.
XX
XX 09-FEB-1998; 98US-0074121.
XX
XX 13-APR-1998; 98US-0081563.

```

```

PR 10-AUG-1998; 98US-0096116.
PR 04-SEP-1998; 98US-0099273.
XX
XX (GEST ) GENSET.
XX
PI Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX
XX WPI: 1999-600966/51.
DR P-PSDB; AA59675.
XX
PT Extended cDNAs useful for expressing secreted proteins and to obtain
PT specific antibodies -
XX
XX Claim 1; Page 168-169; 244pp; English.
XX
CC This sequence encodes a human secreted protein of the invention. The
CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
CC prepare PCR primers and probes. These are useful for forensic matching or
CC positive identification by DNA sequencing. They may also be used in
CC alternative fingerprint identification techniques. Antibodies against the
CC proteins encoded by the extended cDNAs are useful in identification of
CC tissue types or cell species, as well as identifying tissue specific
CC soluble proteins. The sequences can be used for chromosome mapping and
CC identification of genes associated with hereditary diseases or drug
CC response. Signal sequences from the cDNAs can be used in construction of
CC secretion vectors. Other sequences derived from the extended cDNAs can be
CC used to clone upstream genomic DNA sequences including promoters. This is
CC in turn useful for identifying proteins that interact with promoter
CC sequences. Some of the proteins may be useful in diagnosing and treating
CC several disorders including, but not limited to: cancer, hyperlipidaemia,
CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
CC rheumatic diseases, embryonic disorders, hypertension, renal injury,
CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
XX
XX Sequence 1689 BP; 552 A; 350 C; 335 G; 452 T; 0 other;
XX
Query Match          99.7%; Score 321; DB 20; Length 1689;
Best Local Similarity 99.7%; Pred. No. 3.5e-86;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacaagacaacacctgattatcattcatttgatgagcgccacacagtcacg 60
   |||||||
Db 248 gacaagacaacacctgattatcattcatttgatgagcgccacacagtcacg 307
QY 61 tttaagaagaagtgttgcgtaaaataagaataccagaataatggcagagcattgtcct 120
   |||||||
Db 308 tttaagaagaagtgttgcgtaaaataagaataccagaataatggcagagcattgtcct 367
QY 121 cctcaatctggtttatgaacaactgacacacaccttctctcgtatggcagatgtctcc 180
   |||||||
Db 368 cctcaatctggtttatgaacaactgacacacaccttctctcgtatggcagatgtctcc 427
QY 181 cagattatgtttgttaccatctcgcagctgttagagccgatatcattgaagaatattc 240
   |||||||
Db 428 cagattatgtttgttaccatctcgcagctgttagagccgatatcattgaagaatattc 487
QY 241 aaanagctctatgcttaccagacctgcagatcacagctcgtgtgttgcgacaacatgaaga 300
   |||||||
Db 488 aaatgctctatgcttaccagacctgcagatcacagctcgtgtgttgcgacaacatgaaga 547
QY 301 agctctcaagtgtcgtgaagact 322
   |||||||
Db 548 agctctcaagtgtcgtgaagact 569
XX
RESULT 11
ID AAH5526/C
XX
XX AAH5526 standard; DNA; 529 BP.
XX
XX AAH5526;
XX
DT 04-SEP-2001 (first entry)

```

```
XX XX Human breast tumour protein contig 10 DNA sequence.
DE DE
XX XX Cytostatic; vaccine; human; breast tumour protein; breast cancer;
KM KM gene therapy; ds.
XX XX
OS OS Homo sapiens.
XX XX
PN PN WO200140269-A2.
XX XX
PD PD 07-JUN-2001.
XX XX
PF PF 29-NOV-2000; 2000MO-US32520.
XX XX
PR PR 30-NOV-1999; 99US-0451651.
PR PR 22-FEB-2000; 2000US-0510662.
PR PR 10-MAR-2000; 2000US-0523586.
PR PR 07-APR-2000; 2000US-0545068.
PR PR 15-MAY-2000; 2000US-0571025.
XX XX
PA PA (CORI-) CORIXA CORP.
XX XX
PI PI Dillon DC, Day CH, Jiang Y, Houghton RL, Mitcham JL, Wang A;
XX XX WPI; 2001-356154/37.
XX XX
PT PT Breast tumor polypeptides and the nucleic acids that encode them,
XX XX useful for the prevention, diagnosis and treatment of breast cancer -
XX XX
PS PS Claim 24; Page 138-139; 221pp; English.
XX XX
CC CC The present sequence is a human breast tumour protein coding sequence.
CC CC This sequence may be used in the prevention, diagnosis and treatment of
CC CC diseases associated with inappropriate expression of the breast tumour
CC CC protein e.g. breast cancer. For example, this sequence may be used to
CC CC treat disorders associated with decreased expression by rectifying
CC CC mutations or deletions in a patient's genome that affect the activity of
CC CC breast tumour protein by expressing inactive proteins or to supplement
CC CC the patients own production of the breast tumour protein. Additionally,
CC CC the present sequence may be used to produce the breast tumour protein, by
CC CC inserting the nucleic acids into a host cell and culturing the cell to
CC CC express the protein. The present sequence and its complementary sequences
CC CC may also be used as DNA probes in diagnostic assays to detect and
CC CC quantitate the presence of similar nucleic acids in samples, and
CC CC therefore which patients may be in need of restorative therapy.
XX XX
SQ SQ Sequence 529 BP; 147 A; 104 C; 102 G; 176 T; 0 other;

Query Match 96.3%; Score 310; DB 22; Length 529;
Best Local Similarity 99.4%; Pred. No. 4.4e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 gaaacacacacaccccttgatga-ttatcatcacttgatgagtgccacacagtcgaag 59
DB 507 GAAACACACACACCCCTTGATGATTTATTCATCTTGATGATGATGCCACACAGTCAAG 448
QY 60 ctttaagaagaagtggttcggaataaagaatccagaatggcagaacattgtgcc 119
DB 447 CTTTAAAGAAAGTGTTGCGGAAATTAAGAAATCCAGAAATTGGCAGACAGTGTCTCC 388
QY 120 tctcaatcgtgttattgaacaactgacaacaccttctcctgattgagcagatgtcc 179
DB 387 TCCCTCAATCTGTTTATGAAACAACTGACAAACACCTTCTCTGATGGCCAGTATGTCC 328
QY 180 caagattatgttctgtgacccatctctgacagattagagccgatatcaactggaagatatc 239
DB 327 CCAAGGATTAATGTTTGTGACCCATCTCTGACAGTTAGAGCCGATATCACAGGAAGATATT 268
QY 240 caaanngtctctatgcttaagaacctgcagatacagctcgtgtgcttgaacaataaga 299
DB 267 CAAACCGTCTCTATGCTTACGAACCTGCAGATACAGCTCTGTGTTGACACACATGAAGA 208
```

```
QY 300 aagctcgaagttgctgaagact 322
DB 207 AAGCTCTCAAGTTGCTGGAAGACT 185

RESULT 12
ID AAF44884 standard; cDNA; 778 BP.
XX AAF44884;
AC AAF44884;
XX
DT 28-MAR-2001 (first entry)
XX
DE Human breast cancer related protein coding sequence SEQ ID NO: 40.
XX
KM Human; breast cancer; diagnosis; therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200078960-A2.
XX
PD 28-DEC-2000.
XX
PF 23-JUN-2000; 2000MO-US17536.
XX
PR 23-JUN-1999; 99US-0140903.
PR 12-OCT-1999; 99US-0158980.
XX
PA (CORI-) CORIXA CORP.
XX
PI Yuqiu J, Mitcham JL;
XX
DR WPI; 2001-041426/05.
XX
PT New polynucleotides encoding breast tumor specific proteins, useful for
XX PT prevention, treatment and diagnosis of breast cancer -
XX
PS Claim 25; Page 132; 165pp; English.
XX
CC The present invention provides the coding sequences for a number of
CC breast cancer related proteins. These can be used in vaccinations
CC against, diagnosis of and treatment of cancer, particularly breast
CC cancer.
XX
SQ SQ Sequence 778 BP; 245 A; 175 C; 158 G; 193 T; 7 other;

Query Match 96.0%; Score 309; DB 22; Length 778;
Best Local Similarity 99.4%; Pred. No. 1e-82;
Matches 320; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 gacaagcaacaaaccccttgatgattatcatcacttgatgagtgccacacagtcgaagc 60
DB 256 gacaagcaacaaaccccttgatgattatcatcacttgatgagtgccacacagtcgaagc 315
QY 61 tttaagaagaagtggttcggaataaagaatccagaatggcagaacattgttcct 120
DB 316 tttaagaagaagtggttcggaataaagaatccagaatggcagaacattgttcct 375
QY 121 cctcaatcgtgttattgaacaactgacaacaccttctcctgattgagcagatgtccc 180
DB 376 cctcaatcgtgttattgaacaactgacaacaccttctcctgattgagcagatgtccc 435
QY 181 cagattatgttctgtgacccatctctgacagattagagccgatatcaactggaagatatc 240
DB 436 cagattatgttctgtgacccatctctgacagattagagccgatatcaactggaagatatc 495
QY 241 aaanngtctctatgcttaagaacctgcagatacagctcgtgtgcttgaacaataaga 300
DB 496 aaatgctctctatgcttaagaacctgcagat-cagctcgtgtgcttgaacaataaga 554
QY 301 agctcgaagttgctgaagact 322
```

Db 555 agctctcaagttgctgaagact 576

RESULT 13

AAZ40846

ID AAZ40846 standard; DNA: 940 BP.

XX

AC AAZ40846;

XX

DT 18-JAN-2000 (first entry)

XX

DE Secreted protein EST coding sequence 78-21-1-B7-FL1.

XX

KM Secreted protein; fingerprint identification technique;

KM chromosome mapping; human; hereditary disease; diagnosis; cancer;

KM hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;

KM autoimmune disease; rheumatic disease; embryonic disorder; myopathy;

KM renal injury; amino aciduria; hypoglycaemia; male rat infertility;

KM hypertension; ss.

XX

OS Homo sapiens.

XX

PN WO940189-A2.

XX

PD 12-AUG-1999.

XX

PF 09-FEB-1999; 99WO-IB00282.

XX

PR 09-FEB-1998; 98US-0074121.

PR 13-APR-1998; 98US-0081563.

PR 10-AUG-1998; 98US-0096116.

PR 04-SEP-1998; 98US-0099273.

PA

(GEST ) GENSET.

XX

PI Bougueleret L, Duclet A, Dumas Milne Edwards J;

XX

DR WPI: 1999-600966/51.

DR P-PSDB; AAY59718.

XX

PT Extended CDNAS useful for expressing secreted proteins and to obtain

PT specific antibodies -

XX

PS Claim 1: Page 226; 244pp; English.

XX

CC This sequence encodes a human secreted protein of the invention. The

CC extended CDNAS (or genomic DNAs obtainable from them) may be used to

CC prepare PCR primers and probes. These are useful for forensic matching or

CC positive identification by DNA sequencing. They may also be used in

CC alternative fingerprint identification techniques. Antibodies against the

CC proteins encoded by the extended CDNAS are useful in identification of

CC tissue types or cell species, as well as identifying tissue specific

CC soluble proteins. The sequences can be used for chromosome mapping and

CC identification of genes associated with hereditary diseases or drug

CC response. Signal sequences from the CDNAS can be used in construction of

CC secretion vectors. Other sequences derived from the extended CDNAS can be

CC used to clone upstream genomic DNA sequences including promoters. This is

CC in turn useful for identifying proteins that interact with promoter

CC sequences. Some of the proteins may be useful in diagnosing and treating

CC several disorders including, but not limited to: cancer, hyperlipidaemia,

CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and

CC rheumatic diseases, embryonic disorders, hypertension, renal injury,

CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.

XX

SO Sequence 940 BP; 304 A; 202 C; 187 G; 233 T; 14 other;

Query Match 93.9%; Score 302.2; DB 20; Length 940;

Best Local Similarity 96.0%; Pred. No. 1.2e-80;

Matches 310; Conservative 8; Mismatches 4; Indels 1; Gaps 1;

QY 1 gacagcaacaacccctgatattatcatcacttgatgatgagtcacacagtcgaac 60

:|||||

Db 374 racaagcaacaacccctgatattatcatcacttgatgagtcacacagtcgaac 433

QY 61 tttaagaagaagtggttgcgtaaaataagaatccagaattggagagcagttgtct 120

Db 434 tttaaaaaaakgttctcgtaaaaataaataccagaatctggcgtcagttgtccy 493

QY 121 cctcaatctgtttatgaacaactgacaacaccttctcctgtagtgcagtagt-cc 179

Db 494 cctcaatctgtttatgaacaactgacaacaccttctcctgtagtgcagtagtcc 553

QY 180 ccaagattatgtttgttgaccatctctgacagttagagccgatatcaactgaaatatt 239

Db 554 ccmgattatgtttgttgaccatctctgacagttagagccgatatcaactgaaatatt 613

QY 240 caanagctctatgcttgcagacccctgagatagagctgtgtgttgacaacatgaaga 299

Db 614 caaagctctcctaaagcttgcagacccctgagatagagctgtgtgttgacaacatgaaga 673

QY 300 aagctctcaagttgctgaagact 322

Db 674 aagctctcaagttgctgaagact 696

RESULT 14

AAV19188

ID AAV19188 standard; CDNA: 386 BP.

XX

AC AAV19188;

XX

DT 28-AUG-1998 (first entry)

XX

DE Human XAG growth factor huxAG-1 related EST.

XX

KM huxAG-1; XAG: growth factor; colon cancer; tumour marker;

KM breast disease; liver disease; lung disease; emphysema;

KM wound healing; diagnosis; therapy; human; expressed sequence tag;

KW EST; ss.

XX

OS Homo sapiens.

XX

PN WO9807749-A1.

XX

PD 26-FEB-1998.

XX

PF 22-AUG-1997; 97WO-US14139.

XX

PR 23-AUG-1996; 96WO-US13766.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Dillon PJ, Ebner R, Endress GA, Yu G;

XX

DR WPI: 1998-169093/15.

XX

PT New isolated human XAG growth factor(s) - used to develop products

PT for treating e.g. liver, lung or breast diseases or

PT hyperproliferative disorders, e.g. cancer.

XX

PS Claim 5; Page 111; 141pp; English.

XX

CC This expressed sequence tag (EST), deposited as GenBank Accession

CC No. AA244356, is related to a portion of huxAG-1 cDNA (see V19155),

CC which codes for a novel human growth factor (see W37844) that

CC shares homology with the XAG protein of *Xenopus laevis*. Expression

CC of huxAG-1 has been discovered in colon cancer tissue, with no

CC corresponding expression in healthy colon tissue. The huxAG-1 gene

CC provides a molecular marker for colon cancer. Vectors, host cells,

CC antibodies, and screening methods for identifying agonists and

CC antagonists of huxAG-1 are provided. huxAG polypeptides can be

CC used to stimulate proliferation and differentiation of hepatocytes

CC to alleviate or treat liver diseases and pathologies. They can

CC also be used to stimulate or promote liver regeneration, and also

CC to prevent and heal damage to the lungs caused by various

CC pathological states. They can also be used to treat acute or  
CC chronic lung damage, and to stimulate the proliferation and  
CC differentiation of breast tissue.  
XX  
SQ Sequence 386 BP; 122 A; 95 C; 74 G; 95 T; 0 other;

Query Match 85.4%; Score 275; DB 19; Length 386;  
Best Local Similarity 99.6%; Pred. No. 1.1e-72;  
Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaacacccttgatgattatcattcacttgatgagtgccacagtcacg 60  
DB 111 gacaagcaaacacccttgatgattatcattcacttgatgagtgccacagtcacg 170  
QY 61 ttcaagaagaagtgttgcctgaataaagaatccagaatltgcagagcagttgtcct 120  
DB 171 ttcaagaagaagtgttgcctgaataaagaatccagaatltgcagagcagttgtcct 230  
QY 121 cctcaatctggtttatgaacaactgacaacaccccttcctcctgattgcccagtcacg 180  
DB 231 cctcaatctggtttatgaacaactgacaacaccccttcctcctgattgcccagtcacg 290  
QY 181 cagattatgttgcctgagccatctcagacagttagagccgatatcactggaagatctc 240  
DB 291 cagattatgttgcctgagccatctcagacagttagagccgatatcactggaagatctc 350  
QY 241 aaangctctatgcttaagcaacctgcagatacagc 276  
DB 351 aaatgctctatgcttaagcaacctgcagatacagc 386

RESULT 15  
AAF63347  
ID AAF63347 standard; cDNA; 386 BP.

XX AAF63347;

DT 04-MAY-2001 (first entry)

XX HuxAG-1 related EST sequence SEQ ID 35.

XX Human; growth factor; huxAG-1; colon cancer specific gene; CCSG;  
KW cell proliferation; liver disease; fulminant liver failure; cirrhosis;  
KW hepatitis; cancer; colon cancer; colorectal carcinoma; EST;  
KW expressed sequence tag; ss.

OS Homo sapiens.

XX US6171816-B1.

PN 09-JAN-2001.

PD 22-AUG-1997; 97US-0916576.

PF 23-AUG-1996; 96US-0024347.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Yu G, Dillon PJ, Edner R, Endress GA;

PI WPI; 2001-136872/14.

PT Novel human growth factor polypeptide useful for diagnosing and  
PT treating colon cancer and liver diseases, to prevent and heal damage to  
PT the lungs and for identifying modulators of therapeutic use -  
XX

PS Disclosure: Column 79-80; 59pp; English.

XX This invention relates to a human growth factor polypeptide huxAG-1 also  
XX known as a colon cancer specific gene (CCSG). HuxAG-1 stimulates cell  
XX proliferation as a growth factor. The HuxAG-1 protein is useful for  
XX identifying compounds capable of enhancing or inhibiting cellular

CC response induced by huxAG-1. The protein is also useful for stimulating  
CC proliferation of cells e.g. colon, breast, liver and lung cells, and  
CC hepatocytes. It is useful for alleviating or treating liver diseases and  
CC pathologies such as fulminant liver failure caused by cirrhosis, liver  
CC damage caused by viral hepatitis and toxic substances, for preventing and  
CC treating damage to the lungs caused by various pathological states and  
CC for promoting healing of breast tissue injury due to surgery, trauma or  
CC cancer. HuxAG-1 and the identified antagonist are useful for treating  
CC cancer, in particular colon cancer. Detecting altered levels of huxAG-1  
CC and its polynucleotides are useful for diagnosing or detecting cancer in  
CC mammals. The gene encoding huxAG-1 is useful for monitoring human  
CC colorectal carcinoma. huxAG-1 nucleic acid molecules are also useful for  
CC chromosome identification. The present sequence represents an EST  
CC (expressed sequence tag) related to the huxAG-1 sequence.  
CC NOTE: Sequences AAF63347 - AAF63351 are stated as specifically not  
XX claimed in claim 1.

SQ Sequence 386 BP; 122 A; 95 C; 74 G; 95 T; 0 other;

Query Match 85.4%; Score 275; DB 22; Length 386;  
Best Local Similarity 99.6%; Pred. No. 1.1e-72;  
Matches 275; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaacacccttgatgattatcattcacttgatgagtgccacacagtcacg 60  
DB 111 gacaagcaaacacccttgatgattatcattcacttgatgagtgccacacagtcacg 170  
QY 61 tttaagaagaagtgttgcctgaataaagaatccagaatltgcagagcagttgtcct 120  
DB 171 tttaagaagaagtgttgcctgaataaagaatccagaatltgcagagcagttgtcct 230  
QY 121 cctcaatctggtttatgaacaactgacaacaccccttcctcctgattgcccagtcacg 180  
DB 231 cctcaatctggtttatgaacaactgacaacaccccttcctcctgattgcccagtcacg 290  
QY 181 cagattatgttgcctgagccatctcagacagttagagccgatatcactggaagatctc 240  
DB 291 cagattatgttgcctgagccatctcagacagttagagccgatatcactggaagatctc 350  
QY 241 aaangctctatgcttaagcaacctgcagatacagc 276  
DB 351 aaatgctctatgcttaagcaacctgcagatacagc 386

Search completed: March 30, 2002, 13:05:03  
Job time: 21300 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 12:59:16 ; Search time 8456.85 seconds

(without alignments)  
628.141 Million cell updates/sec

Title: US-09-867-034-2

Sequence: 1 gcaacgcaacaaccccttga.....ctctcaagtgtcgaagact 322

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1:  GenBank:
2:  gb_da:
3:  gb_hlg:
4:  gb_in:
5:  gb_ov:
6:  gb_pat:
7:  gb_ph:
8:  gb_pl:
9:  gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_om:
20: em_ov:
21: em_ov:
22: em_pat:
23: em_ph:
24: em_pl:
25: em_ro:
26: em_sts:
27: em_sy:
28: em_un:
29: em_vi:
30: em_htgo_hum:
31: em_htgo_hum:
32: em_htgo_inv:
33: em_htgo_inv:
34: em_htg_inv:
35: em_htg_inv:
36: em_htg_inv:
36: em_htg_inv:

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	99.7	838	9	AF007791 Homo sapi
2	321	99.7	875	6	AR123828 Sequence
3	321	99.7	956	6	AF088867 Homo sapi
4	321	99.7	1020	6	AX011612 Sequence
5	321	99.7	1020	6	AX015056 Sequence
6	321	99.7	1077	6	AF038451 Homo sapi
7	310	96.3	529	6	AX156194 Sequence
8	309	96.0	778	6	AX067336 Sequence
9	275	85.4	386	6	AR123857 Sequence
10	263.4	81.5	401	6	AX192465 Sequence
11	262.4	81.5	758	10	AB014592 Sequence
12	262.4	81.5	781	10	AF044262 Sequence
13	249	77.3	409	6	AX192616 Sequence
14	246.8	76.6	793	6	AR144213 Sequence
15	205.8	63.9	506	6	AR123832 Sequence
16	159.4	49.5	235	6	AX062783 Sequence
17	159.4	49.5	301	6	AX192553 Sequence
18	157.2	48.8	640	6	AX193063 Sequence
19	157.2	48.8	732	6	AR123830 Sequence
20	149.4	46.4	420	6	AX098187 Sequence
21	136	42.2	462	5	AF025474 Sequence
22	133.8	41.6	1004	5	AF025474 Sequence
23	132.8	41.2	1946	5	AXU82110 Sequence
24	125.8	39.1	2057	5	AXU76752 Sequence
25	95.8	29.8	401	6	AX193073 Sequence
26	95.8	29.8	410	6	AX193110 Sequence
27	93.4	29.0	489	6	AR123831 Sequence
28	90.2	28.0	160274	2	AC073411 Homo sapi
29	90.2	28.0	165616	9	AC073333 Homo sapi
30	90.2	28.0	169554	2	AC021379 Homo sapi
31	76.6	23.8	169554	2	AC021379 Homo sapi
32	59	18.3	329	6	AR123833 Sequence
33	59	18.3	329	6	AR123834 Sequence
34	59	18.3	329	6	AR123835 Sequence
35	59	18.3	329	6	AR123836 Sequence
36	57.8	18.0	330	6	AR123837 Sequence
37	53.6	16.6	1315	10	BC006857 Mus muscu
38	46	14.3	640	6	AX136502 Sequence
39	45.8	14.2	1391	6	AX136177 Sequence
40	45.8	14.2	1409	9	BC008953 Homo sapi
41	45.8	14.2	1422	9	BC001493 Homo sapi
42	45.8	14.2	1423	6	AR123829 Sequence
43	45.8	14.2	1452	9	AR131758 Homo sapi
44	44.2	13.7	1450	9	BC008913 Homo sapi
45	43	13.4	311	6	AR123862 Sequence

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE
1	AF007791	838 bp mRNA	AF007791	AF007791	GI:3779196	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	Thompson, D.A. and Weigel, R.J.	hag-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines	Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)	99009231	2 (bases 1 to 838)

AUTHORS Thompson, D.A.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUN-1997) Surgery, Stanford, 1201 Welch Road, MSLs  
Building, Room P228, Stanford, CA 94305, USA  
FEATURES Location/Qualifiers  
source 1..838  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue.type="breast"  
/cell\_line="adenocarcinoma"  
/cell\_line="MCF7; estrogen receptor-expressing"  
1..838  
/gene="hag-2/c"  
11..538  
/note="similar to Homo sapiens protein hag-2/R deposited in GenBank Accession Number AF038451, and to Xenopus laevis cement gland proteins np77 deposited in GenBank Accession Number U82110 and XAG-2 deposited in GenBank Accession Number AF025474"  
/codon\_start=1  
/product="secreted cement gland protein XAG-2 homolog"  
/protein\_id="AAC7358.1"  
/db\_xref="GI:3779197"  
/translation="MEKIPVSAFLILVALSYTLARDFTVYKPGAKKDKSRPKLPQTL SRMGDQLIMTQTYEALYKSKTSNKLPMI IHLLDECPHSAALKVFAENKEIOKLAE OFVLNLVETTDKHLSPDGQYVPRIMFVDPSTLVADITGRYSNRLVAYEPADTALL LNMKKALKLKTTEL"

BASE COUNT 291 a 170 c 160 g 217 t  
ORIGIN

Query Match 99.7%; Score 321; DB 9; Length 838;  
Best Local Similarity 99.7%; Pred. No. 3.1e-83;

Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaacaacccttgatgattatcattcacttgatgagtgccacacagtcgaac 60  
|||||  
DB 208 GACAAGCAACAACCCTTGATGATTTATTCACCTTGATGAGTGGCCACACAGTCAAGC 267  
QY 61 tttaagaagaagtgttgcctaaataaagaatccagaatcttgacagagttgtcct 120  
|||||  
DB 268 TTTAAGAAGAAGTGTGCTAAATAAGAAATCCAGAAATGGCAGAGAGTTGTGCT 327  
QY 121 cctcaatctgtttatgaaacaactgacaacaccttctcctatgagcagatgtcc 180  
|||||  
DB 328 CCTCAATCTGCTTATTAAGCAACACCTTTCTCCTGATGCGCAGATATGCC 387  
QY 181 cagattatgtttgttgagccatctctgacagttagagccgatatcaactggaagata 240  
|||||  
DB 388 CAGGATTATGTTGTTGACCCATCTCGACAGTTAGCCGATATTCACCTGGAAGATATTC 447  
QY 241 aaangctctatgctgaagaacctgagatagagctctgtgcttacaacatgaagaa 300  
|||||  
DB 448 AAATGCTCTTATGCTTACGACACCTGAGATACAGCTCTGTGCTTACACATGAAGAA 507  
QY 301 agctctcaagttgctgaagact 322  
|||||  
DB 508 AGCTCTCAAGTTGCTGAAGACT 529

RESULT 2  
LOCUS ARI23828 875 bp DNA PAT 16-MAY-2001  
DEFINITION Sequence 1 from patent US 6171816.  
ACCESSION ARI23828  
VERSION ARI23828.1 GI:14109189  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 875)  
AUTHORS Yu, G., Dillon, P.J., Ebner, R. and Endress, G.A.

TITLE Human XAG-1 polynucleotides and polypeptides  
JOURNAL Patent: US 6171816-A 1 09-JAN-2001;  
FEATURES Location/Qualifiers  
source 1..875  
/organism="unknown"  
BASE COUNT 283 a 189 c 180 g 223 t  
ORIGIN

Query Match 99.7%; Score 321; DB 6; Length 875;  
Best Local Similarity 99.7%; Pred. No. 3.1e-83;

Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaacaacccttgatgattatcattcacttgatgagtgccacacagtcgaac 60  
|||||  
DB 268 GACAAGCAACAACCCTTGATGATTTATTCACCTTGATGAGTGGCCACACAGTCAAGC 327  
QY 61 tttaagaagaagtgttgcctaaataaagaatccagaatcttgacagagttgtcct 120  
|||||  
DB 328 TTTAAGAAGAAGTGTGCTAAATAAGAAATCCAGAAATGGCAGAGAGTTGTGCT 387  
QY 121 cctcaatctgtttatgaaacaactgacaacaccttctcctatgagcagatgtcc 180  
|||||  
DB 388 CCTCAATCTGCTTATTAAGCAACACCTTTCTCCTGATGCGCAGATATGCC 447  
QY 181 cagattatgtttgttgagccatctctgacagttagagccgatatcaactggaagata 240  
|||||  
DB 448 CAGGATTATGTTGTTGACCCATCTCGACAGTTAGCCGATATTCACCTGGAAGATATTC 507  
QY 241 aaangctctatgctgaagaacctgagatagagctctgtgcttacaacatgaagaa 300  
|||||  
DB 508 AAATGCTCTTATGCTTACGACACCTGAGATACAGCTCTGTGCTTACACATGAAGAA 567  
QY 301 agctctcaagttgctgaagact 322  
|||||  
DB 568 AGCTCTCAAGTTGCTGAAGACT 589

RESULT 3  
LOCUS AF088867 956 bp mRNA PRI 02-JAN-2000  
DEFINITION Homo sapiens putative secreted protein XAG mRNA, complete cds.  
ACCESSION AF088867  
VERSION AF088867.1 GI:6652811  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 956)  
AUTHORS Zhang, J.S. and Smith, D.I.  
TITLE Human homolog of XAG is differentially expressed in tumors  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 956)  
AUTHORS Zhang, J.S. and Smith, D.I.  
TITLE Direct Submission  
JOURNAL Submitted (30-AUG-1998) Pathology and Lab. Medicine, Mayo Clinic, 200 SW 1st St., Rochester, MN 55905, USA  
FEATURES Location/Qualifiers  
source 1..956  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
148..675  
/codon\_start=1  
/product="putative secreted protein XAG"  
/protein\_id="AAF22484.1"  
/db\_xref="GI:6652812"  
/translation="MEKIPVSAFLILVALSYTLARDFTVYKPGAKKDKSRPKLPQTL SRMGDQLIMTQTYEALYKSKTSNKLPMI IHLLDECPHSAALKVFAENKEIOKLAE OFVLNLVETTDKHLSPDGQYVPRIMFVDPSTLVADITGRYSNRLVAYEPADTALL LNMKKALKLKTTEL"

BASE COUNT 310 a 202 c 202 g 242 t  
ORIGIN



Query Match 99.7%; Score 321; DB 9; Length 956;  
Best Local Similarity 99.7%; Pred. No. 3.le-83;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagcaacaaccccttgatgatcattatcattcattgagtgagccacacagtcacg 60  
DB 345 GACAGCAACAACCCCTTGATGATATTATTCATCAGTGTGATGAGCCACACAGCAAGC 404

QY 61 tttaagaagaagtgcttcgtaaaataaagaatccagaatctggcagagcagttgtcct 120  
DB 405 TTTAAGAAGAAGTGTGCTGAAAATAAGAAATCCAGAAATGGCAGAGCATGTGTCTCT 464

QY 121 cctcaatcgtttatgaataaactgaacaccccttcctcattatggccagttatgtcc 180  
DB 465 CCTCAATCTGTTATGTAAGAACACTGACAAACACCTTCTCTGATGAGCCAGTATGTCC 524

QY 181 cagattatgttctgacccatctctgacagttagagccgatalcaactggagaatattc 240  
DB 525 CAGATTATGTTGTTGTGACCATCTCTGACAGTTAGAGCCGATATTCAGTGAAGTATTC 584

QY 241 aaangctctctatgcttgaagaacctgcagatagacgctctgttgccttgacaacatgaaga 300  
DB 585 AAACGCTCTATGCTTACGAACTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAA 644

QY 301 agcttcaagttgctgaagact 322  
DB 645 AGCTCTCAAGTTGCTGGAAGACT 666

RESULT 4  
AX011612 1020 bp DNA PAT 06-SEP-2000  
LOCUS Sequence 8 from Patent WO9955858.  
DEFINITION AX011612  
ACCESSION AX011612.1 GI:9998136  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1020)  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pillarsky,C.  
TITLE Human nucleic acid sequences obtained from pancreas tumor tissue  
JOURNAL Patent: WO 9955858-A 8 04-NOV-1999;  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILLARSKY CHRISTIAN (DE)  
FEATURES  
source 1..1020  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 341 a 214 c 218 g 247 t  
ORIGIN

Query Match 99.7%; Score 321; DB 6; Length 1020;  
Best Local Similarity 99.7%; Pred. No. 3.le-83;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 181 cagattatgttcttgacccatctctgacagttagagccgatatcactggaagatattc 240  
DB 527 CAGATTATGTTGTTGTTGACCATCTCTGACAGTTAGAGCCGATATTCAGTGAAGTATTC 586

QY 241 aaangctctctatgcttgcagaacctgcagatagacgctctgttgccttgacaacatgaaga 300  
DB 587 AAACGCTCTATGTTATGTAAGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAA 646

QY 301 agcttcaagttgctgaagact 322  
DB 647 AGCTCTCAAGTTGCTGGAAGACT 668

RESULT 5  
AX015056 1020 bp DNA PAT 07-SEP-2000  
LOCUS Sequence 265 from Patent WO9953040.  
DEFINITION AX015056  
ACCESSION AX015056  
VERSION AX015056.1 GI:10041195  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1020)  
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
Pillarsky,C.  
TITLE Human nucleic acid sequences from ovarian tumour tissue  
JOURNAL Patent: WO 9953040-A 265 21-OCT-1999;  
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
(DE); PILLARSKY CHRISTIAN (DE)  
FEATURES  
source 1..1020  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 341 a 214 c 218 g 247 t  
ORIGIN

Query Match 99.7%; Score 321; DB 6; Length 1020;  
Best Local Similarity 99.7%; Pred. No. 3.le-83;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacagcaacaaccccttgatgatcattatcattcattgagtgagccacacagtcacg 60  
DB 347 GACAGCAACAACCCCTTGATGATATTATTCATCAGTGTGATGAGCCACACAGCAAGC 406

QY 61 tttaagaagaagtgcttcgtaaaataaagaatccagaatctggcagagcagttgtcct 120  
DB 407 TTTAAGAAGAAGTGTGCTGAAAATAAGAAATCCAGAAATGGCAGAGCATGTGTCTCT 466

QY 121 cctcaatcgtttatgaataaactgaacaccccttcctcattatggccagttatgtcc 180  
DB 467 CCTCAATCTGTTATGTAAGAACACTGACAAACACCTTCTCTGATGAGCCAGTATGTCC 526

QY 181 cagattatgttctgacccatctctgacagttagagccgatalcaactggagaatattc 240  
DB 527 CAGATTATGTTGTTGTGACCATCTCTGACAGTTAGAGCCGATATTCAGTGAAGTATTC 586

QY 241 aaangctctctatgcttgcagaacctgcagatagacgctctgttgccttgacaacatgaaga 300  
DB 587 AAACGCTCTATGCTTATGTAAGAACCTGCAGATACAGCTCTGTTGCTTGACAACATGAAGAA 646

QY 301 agcttcaagttgctgaagact 322  
DB 647 AGCTCTCAAGTTGCTGGAAGACT 668

RESULT 6  
AF038451 1077 bp mRNA PRI 03-DEC-1998  
LOCUS

```

DEFINITION Homo sapiens secreted cement gland protein XAG-2 homolog (hag-2/R)
ACCESSION AF038451
VERSION AF038451.1 GI:3779225
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1077)
AUTHORS Thompson,D.A. and Weigel,R.J.
TITLE hag-2, the human homologue of the Xenopus laevis cement gland gene
XAG-2, is coexpressed with estrogen receptor in breast cancer cell
lines
JOURNAL Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)
MEDLINE 99009231
REFERENCE 2 (bases 1 to 1077)
AUTHORS Thompson,D.A.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1997) Surgery, Stanford University, 1201 Welch
Road, MSLS Building, Room P228, Stanford, CA 94305, USA
FEATURES
source
1.1077
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="breast"
/cell_line="MCF7"
/note="Isolated from estrogen receptor-positive breast
adenocarcinoma cell line"
1.1077
/gene="hag-2/R"
59..586
/note="similar to Homo sapiens protein hag-2/C deposited
under GenBank Accession Number AF007791, and to Xenopus
laevis cement gland-specific proteins np77 and XAG-2
deposited under GenBank Accession Numbers U82110 and
AF025474, respectively"
/codon_start=1
/product="secreted cement gland protein XAG-2 homolog"
/protein_id="AAC82614.1"
/db_xref="GI:3779226"
/translacion="MEKIPVSAFLLLVALSTYLANDTVKPGAKKDKDSRPRLPQTL
SRGWDOLIMTQTEBALYKSKTSNKPIMIHLDECPHSAKLVFAKNKELQIAE
QFVLNLVEYTDKHLSPDGOYVPRIMEFVPSLTVRADITGRSNRLYAYEPADVAL
LDNMKKALKLKTLEL"
BASE COUNT 367 a 231 c 210 g 269 t
ORIGIN
Query Match 99.7%; Score 321; DB 9; Length 1077;
Best Local Similarity 99.7%; Pred. No. 3.2e-83;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 gacaaacaacacccctgatgatattcaatcacttgatgagtgccacacagtcgaacg 60
DB 256 GACAAACCAACAAACCCCTGATGATTAATTCATTGAGTGAAGTCCACACAGTCAAGC 315
QY 61 tttaagaagaagtgctgcgaataaagaatccaagaattgcagagagattgtcct 120
DB 316 TTTAAAGAAAGTGTGCTGAATAAAGAAATCCAGAAATTCGACAGAGATTGTGCTT 375
QY 121 cctcaatcgtgtttatgaacaacactgacaacaccccttcctcgcgaatggcagatgtccc 180
DB 376 CCTCATCTGCTGTTATGAACAACTGACAAACCTTTCTCCTGATGCGCAGATATGCC 435
QY 181 caggaattatgttcttgaccatctcctgacagttagagccgatatcaactggaatatlc 240
DB 436 CAGGATTATGTTGTTGACCCATCTCGACAGTTAGAGCCGATATCAGTGAAGATATTC 495
QY 241 aaangctctatgcttaagaacctgcagataagctctgttcttgacaaactgaagaa 300
DB 496 AATCTCTCTATGCTTACGAAACCTGCAGATACAGCTCTGTTGCTTGACAAACATGAAGA 555

```

```

QY 301 agctctcaagttgctgaagact 322
DB 556 AGCTCTCAAGTTGCTGAAGACT 577
RESULT 7
AX156194/c AX156194 529 bp DNA PAT 22-JUN-2001
LOCUS Sequence 51 from Patent WO0140269.
DEFINITION AX156194
ACCESSION AX156194
VERSION AX156194.1 GI:14537195
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 529)
AUTHORS Dillon,D.C., Day,C.H., Jiang,Y., Houghton,R.L., Mitcham,J.L. and
Wang,A.
TITLE Compositions and methods for therapy and diagnosis of breast cancer
JOURNAL Patent: WO 0140269-A 51 07-JUN-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 147 a 104 c 102 g 176 t
ORIGIN
Query Match 96.3%; Score 310; DB 6; Length 529;
Best Local Similarity 99.4%; Pred. No. 4.8e-80;
Matches 321; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 1 gacaaacaacaccccttgatga-ttatcacaacttgatgagtgccacacagtcgaag 59
DB 507 GACAAACCAACAAACCCCTGATGATTAATTCATTGATGATGAGTCCACACAGTCAAG 448
QY 60 cttaagaagaagtgctgcgaataaagaatccaagaattgcagagagattgtcct 119
DB 447 CTTTAAAGAAAGTGTGCTGCAAAATTAAGAAATCCAGAAATTCGACAGCACTTTGTCC 388
QY 120 tctcaatcgtgtttatgaacaacactgacaacaccccttcctcctgaatggcagatgtcc 179
DB 387 TCTCATATCTGCTTATGAACAACTGACAAACACCTTCTCCTGATGCGCAGTATGTCC 328
QY 180 ccaagattatgttcttgaccatctcctgacagttagagccgatatcaactggaatatlc 229
DB 327 CCAGGATTATGTTGTTGACCCATCTCTGACAGTTAGAGCCGATATCAGTGAAGATATT 268
QY 240 caaangctctatgcttgaagaacctgcagatatcaagctctgttcttgacaaactgaagaa 299
DB 267 CAACCGTCTCTATGCTTACGAAACCTGCAGATACAGCTCTGTTGCTTGACAAACATGAAGA 208
QY 300 aagctctcaagttgctgaagact 322
DB 207 AAGCTCTCAAGTTGCTGAAGACT 185
RESULT 8
AX067336 AX067336 778 bp DNA PAT 24-JAN-2001
LOCUS Sequence 40 from Patent WO0078960.
DEFINITION AX067336
ACCESSION AX067336
VERSION AX067336.1 GI:12544960
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 778)
AUTHORS Yudin,J. and Mitcham,J.L.

```

TITLE	Compositions and methods for the therapy and diagnosis of breast cancer			
JOURNAL	Patent: WO 0078960-A 40 28-DEC-2000;			
FEATURES	CORIXA CORPORATION (US)			
source	Location/Qualifiers			
	1. .778			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	1. .778			
misc.feature	/note="n = A,T,C or G"			
	1. .778			
BASE COUNT	245 a	175 c	158 g	193 t
ORIGIN	7 others			
Query Match	96.0%; Score 309; DB 6; Length 778;			
Best Local Similarity	99.4%; Pred. No. 9.8e-80;			
Matches 320;	Conservative	0;	Mismatches	1; Indels 1; Gaps 1;
OY	1	gacaaagcaacaaccccttgatgatattatcatcactctgtgatgagtgccacacagtcgaagc	60	
Db	256	GACAAAGCAACAACCCCTTGATGATATTATTCATCATCTGGATGAGTGCCACACAGCAAGC	315	
OY	61	tttaaagaagaattgtttgttgtaaaaataaagaataatccaagaatgtgcaagagcagttgtcct	120	
Db	316	TTTAAAGAAAGTGTTGTTGCTGTAATAAAGAAATCCAGAAATTTGGCAGAGCAGATTGTCTCT	375	
OY	121	cctcaatctgtgtttaaataacacactgcagacaacaccttctcctgtatgtgcagatgtccc	180	
Db	376	CCTCAATCTGGTTTATGAAACACTGACAAACACCTTTCTCTGTATGGCAGATATGTCCC	435	
OY	181	cagagattatgtttgttgtaaccatctctgcacagttgagagccgatatacactgtgaagatattc	240	
Db	436	CAGATTATGTGTTGTGACCCCATCTCAGACATTGAGACCCGATATACACTGGAAATATTTC	495	
OY	241	aaanqctctcatgcttcgaacctgcagacagatacagctctgtgttgctgaacaataagaa	300	
Db	486	AAATGCTCTATGCTTATGCAAGCACTGCGAGAT-CAGCTGTGTTGCTTGGACACATGAGAA	554	
OY	301	agctctcaagttgctgcgaagact	322	
Db	555	AGCTCTCAAGTTGCTGGAAGACT	576	
RESULT	9			
LOCUS	ARI23857	386 bp	DNA	PAT
DEFINITION	Sequence 35 from patent US 6171816.			
ACCESSION	ARI23857			16-MAY-2001
VERSION	ARI23857.1	GI:14109218		
KEYWORDS	Unknown.			
SOURCE	Unknown.			
ORGANISM	Unclassified.			
AUTHORS	1 (bases 1 to 386)			
TITLE	Yu,G., Dillon,P.J., Ebner,R. and Endress,G.A.			
JOURNAL	Human XAG-1 polynucleotides and polypeptides			
FEATURES	Patent: US 6171816-A 35 09-JAN-2001;			
source	Location/Qualifiers			
	1. .386			
BASE COUNT	122 a	95 c	74 g	95 t
ORIGIN	/organism="unknown"			
Query Match	85.4%; Score 275; DB 6; Length 386;			
Best Local Similarity	99.6%; Pred. No. 7.9e-70;			
Matches 275;	Conservative	0;	Mismatches	1; Indels 0; Gaps 0;
OY	1	gacaagcaacaaccccttgatgatattatcatcactctgtgatgagtgccacacagtcgaagc	60	
Db	111	GACAAGCAACAACCCCTTGATGATATTATTCATCATCTGGATGAGTGCCACACAGTCAAGC	170	
OY	61	tttaaagaagaattgtttgttgtaaaaataaagaataatccaagaatgtgcaagagcagttgtcct	120	

Db	171	TTTAAAGAAAGTCTTGGCTGAAAATAAAGAAATCCAGAAATTGGACAGACAGTTTGCTCT	230
Qy	121	cccaatcctggtttatgtgaacaacacgtgaacaacaccttctctctgtagtcagatgltccc	180
Db	231	CCTCAATCGTGGTTTATGAAACAACATGACAAACACCTTTCTCTGTATGGCCAGTATGTC	290
Qy	181	caggataltglttgttgaccatctctcagatgtagagccgatalcactggaagatltc	240
Db	291	CAGGATTAATGTTGTTGTGACCCATCTCTACAGTTTGTAGACCCGATATCACTGGAAATATT	350
Qy	241	aaacgtctctatgcttgcgaacccgcgagatacgc	276
Db	351	AAATGCTCTTATGCTTACGAACCTGCAGATACAGC	386
RESULT 10			
AX192465	AX192465	401 bp	DNA
LOCUS	Sequence 32 from Patent WO0149716.		PAT
DEFINITION	AX192465		15-AUG-2001
ACCESSION	AX192465		
VERSION	AX192465.1	GI:15210429	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 401)		
JOURNAL	Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,		
FEATURES	Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.		
SOURCE	Compounds for immunotherapy and diagnosis of colon cancer and methods for their use		
LOCATION/Qualifiers	Patent: WO 0149716-A 32 12-JUL-2001;		
1:401	CORIXA CORPORATION (US)		
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
BASE COUNT	128 a	101 c	77 g
ORIGIN			95 t
Query Match	81.8%;	Score 263.4;	DB 6;
Best Local Similarity	98.9%;	Pred. No. 1.9e-66;	
Matches 275; Conservative	0;	Mismatches 2;	Indels 1;
Gaps 1;			
Qy	1	gacgaagcaaacacccatgatcatatcatcacttgtagtgcgcaacagccaagc	60
Db	124	GACAAAGCAACACCCCTGATGATATTCATCTTGGGTGAGTCCACACAGTCAAGC	183
Qy	61	tttaagaagaagtgcttgcgtgaataataagaatccagaatgycagagcaattgctcct	120
Db	184	TTTAAAGAAAGTCTTGCAGAAATAAATCCAGAAATTGGCAGAGCAATTTGTCT	243
Qy	121	cctcaatcgtgttattgnaacaactgcaaacaccccttctcctgtagtgcagatgccc	180
Db	244	CCTCAATCGTGGTTTATGAAACAACATGACAAACACCTTCTCTGATGGCCAGTATGTC	303
Qy	181	caggataltgltgttgaccatctctcagatgtagag-cgagatcactggaagatlt	239
Db	304	CAGGATTAATGTTGTTGTGACCCATCTCTGACATGTAGAGCCGATATCACTGGAAATATT	363
Qy	240	caaacgtctctatgcttgcgaacccgcgagatacgc	277
Db	364	CAACCGTCTCTATGCTTACGAACCTGCAGATACAGCT	401
RESULT 11			
LOCUS	AB016592	758 bp	mRNA
DEFINITION	Mus musculus mRNA for G0B-4, complete cds.		ROD
ACCESSION	AB016592		10-NOV-1999
VERSION	AB016592.1	GI:3395165	

KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

GOB-4.  
Mus musculus (strain:Balb/c) adult intestine goblet cell cDNA to mRNA, clone:GOB-4.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (sites)  
Komiya,T., Tanigawa,Y. and Hirohashi,S.  
Cloning of the gene gob-4, which is expressed in intestinal goblet cells in mice  
Biochim. Biophys. Acta 1444 (3), 434-438 (1999)  
99196718  
2 (bases 1 to 758)  
Komiya,T. and Hirohashi,S.  
Direct Submission  
Submitted (29-JUL-1998) to the DDBJ/EMBL/Genbank databases. Tohru Komiya, ERATO, JST, Genome Asymmetry Group, Doi Biosymmetry Project, Tokodai 5-9-9, Tsukuba, Ibaragi 300-2635, Japan  
(E-mail:tkom@bioa.eraso.trc-net.co.jp, Tel:81-298-48-1515, Fax:81 298-47-8901)

FEATURES  
source  
location/Qualifiers  
1..758  
/organism="Mus musculus"  
/strain="Balb/c"  
/db\_xref="taxon:10090"  
/cell\_type="goblet cell"  
/clone="Gob-4"  
/dev\_stage="adult"  
/tissue\_type="intestine"  
52..579  
/gene="gob-4"  
52..579  
/gene="gob-4"  
/codon\_start=1  
/product="GOB-4"  
/protein\_id="BA32044.1"  
/db\_xref="GI:3395166"  
/translation="MEKFSVSAILLVAISGLAKDTYVSGAKKDPKDSRPRLPQTL SRMGDQLIWTOYVEALYRSKTSNRPILVHHLDECPHSOALKVFAEHKEIOKLAE QEVLLNLVYETDVKHLSPDGQYVPRIVFVPSLTVRADITGRYSNRLYAVEPSDIAL YDNMKKALKLKTLEL"  
758  
polyA\_site  
/note="18 a nucleotides"  
BASE COUNT 234 a 185 c 165 g 174 t  
ORIGIN

Query Match 81.5%; Score 262.4; DB 10; Length 758;  
Best Local Similarity 88.5%; Pred. No. 4.1e-66;  
Matches 284; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 gacaaagcaaacacccctgatgatattcatcacttgatgagtgccacacagtcgaagc 60  
|||||  
DB 249 GACACAGCAACAGACCCCTTGATGTCATTCATCTTGAGAGAAATGCCACACATCAAGC 308  
|||||  
QY 61 ttcaagaagaagtgttgctgtaaaataagaataatccagaatctggcagagcagttgtcct 120  
|||||  
DB 309 CTTAAAGAAAGTGTTCCTGACATTAAGAAATCCAGAAATGGCAGAGAGTTGTTCT 368  
|||||  
QY 121 cctcaatctggtttatgaacaactgacaacaccccttcctcctgagtgccagtatgtccc 180  
|||||  
DB 369 CCTCAACCTGCTATGTAACACCGCAACGACCTTTCTCTGATGGCCAGTACGTC 428  
|||||  
QY 181 caagatatgtttgtagccatctctgcagcttagacgcatatcactggaagatattc 240  
|||||  
DB 429 CAAATTTGTTTGTATGACACCATCCCTGACGGTGAGGGCGACATCTCTGGACATCTC 488  
|||||  
QY 241 aaanagctctatagctatgacgaactgcagataagagctctgttgcttgaacatagaagaa 300  
|||||  
DB 489 AAACGGGCTCTAGAGCTTATGTAACCTTCTGACACAGCTTTGTTGACGACAACTGAAGAA 548  
|||||  
QY 301 agctctcaagttgctctgaagac 321  
|||||

Db 549 AGCTTCACAGCTGCTAAAGAC 569

RESULT 12  
AF044262  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

AF044262 781 bp mRNA ROD 11-NOV-1998  
Mus musculus, secreted cement gland protein XAG-2 homolog (mAg-2)  
mRNA, complete cds.  
AF044262.1 GI:3779231  
house mouse.  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 781)  
Thompson,D.A. and Weigel,R.J.  
hAg-2, the human homologue of the Xenopus laevis cement gland gene XAG-2, is coexpressed with estrogen receptor in breast cancer cell lines  
Biochem. Biophys. Res. Commun. 251 (1), 111-116 (1998)  
9909231  
2 (bases 1 to 781)  
Thompson,D.A.  
Direct Submission  
Submitted (22-JAN-1998) Surgery, Stanford University, MSLS Building, Room P228, 1201 Welch Road, Stanford, CA 94305, USA

FEATURES  
source  
location/Qualifiers  
1..781  
/organism="Mus musculus"  
/strain="Swiss Webster"  
/db\_xref="taxon:10090"  
/dev\_stage="17-day embryo"  
1..781  
/gene="mAg-2"  
53..580  
/note="similar to Homo sapiens proteins hAg-2/R deposited under GenBank Accession Number AF038451 and hAg-2/C deposited under GenBank Accession Number AF007791, and to Xenopus laevis cement-gland proteins m977 deposited under GenBank Accession Number U82110 and XAG-2 deposited under GenBank Accession Number AF025474"  
/codon\_start=1  
/product="secreted cement gland protein XAG-2 homolog"  
/protein\_id="AAC72705.1"  
/db\_xref="GI:3779232"  
/translation="MEKFSVSAILLVAISGLAKDTYVSGAKKDPKDSRPRLPQTL SRMGDQLIWTOYVEALYRSKTSNRPILVHHLDECPHSOALKVFAEHKEIOKLAE QEVLLNLVYETDVKHLSPDGQYVPRIVFVPSLTVRADITGRYSNRLYAVEPSDIAL YDNMKKALKLKTLEL"  
781  
BASE COUNT 256 a 185 c 166 g 174 t  
ORIGIN

Query Match 81.5%; Score 262.4; DB 10; Length 781;  
Best Local Similarity 88.5%; Pred. No. 4.1e-66;  
Matches 284; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 gacaaagcaaacacccctgatgatattcatcacttgatgagtgccacacagtcgaagc 60  
|||||  
DB 250 GACACAGCAACAGACCCCTTGATGTCATTCATCTTGAGAGAAATGCCACACATCAAGC 309  
|||||  
QY 61 tttaagaagaagtgttgctgtaaaataagaataatccagaatctggcagagcagttgtcct 120  
|||||  
DB 310 CTTAAAGAAAGTGTTCCTGACATTAAGAAATCCAGAAATGGCAGAGAGTTGTTCT 369  
|||||  
QY 121 cctcaatctggtttatgaacaactgacaacaccccttcctcctgagtgccagtatgtccc 180  
|||||  
DB 370 CCTCAACCTGCTATGTAACACCGCAACGACCTTTCTCTGATGGCCAGTACGTC 429  
|||||  
QY 181 caagatatgtttgtagccatctctgcagcttagacgcatatcactggaagatattc 240  
|||||

Db	430	CAGAAATTGGTTGTTTGAAGAACCCATCCCTGACGGTGAAGGGACAGACATACACATCGACGATATAC	489
Qy	241	aaancgtctctatgcttctcgaacctgcagatatacagctctgttgccttgacacatgaaga	300
Db	430	AAACGGGCTTACGCTTATGAAACCTTCTGACACAGCTTGTGTACGACACATGAGAA	549
Qy	301	agctctcaagtgtcgtgaagac	321
Db	550	AGCTCTCAAGCTGCTAAAGAC	570
RESULT	13		
LOCUS	AX192616		
DEFINITION	Sequence 183 from Patent WO0149716.	PAT	15-AUG-2001
ACCESSION	AX192616		
VERSION	AX192616.1		
KEYWORDS	GI:15210579		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	EnxarVota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J., Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.		
TITLE	Compounds for immunotherapy and diagnosis of colon cancer and methods for their use		
JOURNAL	Patent: WO 0149716-A 183 12-JUL-2001;		
FEATURES	CORIXA CORPORATION (US)		
SOURCE	location/qualifiers		
BASE COUNT	131 a 104 c 80 g 94 t		
ORIGIN	1		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
Query Match	77.3%;	Score 249;	DB 6;
Best Local Similarity	99.6%;	Pred. No. 3.2e-62;	
Matches 249;	Conservative	0;	Mismatches 1;
Indels	0;	Gaps	0;
Qy	1	gacaaagcacaacaccccttgatgatattatcattacttgatggtgcgccacacagtcacagc	60
Db	160	GACAAAGCAACAAACCCCTTGATGATTTATTCATCATCTTGATGAGTGCCACACAGTCAAGC	219
Qy	61	tttaagaagaagtgttctgcgtgaataataagaataccaagaatgagcagagcagttgtcct	120
Db	220	TTTAAAGAAAGCTTGTTCGGAATAATAAACAATCCAGAAATTGGCAGACAGTTGTCTCT	279
Qy	121	cctcaatcgtgttattgaacaacactgacaacacacttctcctgtatggccagtatgtccc	180
Db	280	CCTCAATCTGCTTATGAAACAACAGACAAACACCTTCTCCTGATGGCCAGTATGTCCC	339
Qy	181	caggatctgttctgtgaccccatctctcagcagttgagccagatatacctcgtgaagatattc	240
Db	340	CAGGATATAGTTGTTGTTGACCCATCTCTGTGACAGTTAGAGCCGATATTCACGAAAGATATTC	399
Qy	241	aaancgtctc	250
Db	400	AAATGCTCTC	409
RESULT	14		
LOCUS	ARI44213		
DEFINITION	Sequence 78 from patent US 6210883.	PAT	08-AUG-2001
ACCESSION	ARI44213		
VERSION	ARI44213.1		
KEYWORDS	GI:15106080		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 793)		

FEATURES	source	1..793	/organism="unknown"
BASE COUNT	246 a	172 c	148 g
ORIGIN			211 t
			16 others
Query Match	Best Local Similarity	76.6%;	Score 246.8; DB 6; Length 793;
	Matches 299; Conservative	0;	Mismatches 16; Indels 5; Gaps 4;
FEATURES			
source			
BASE COUNT	246 a	172 c	148 g
ORIGIN			211 t
			16 others
Query Match	Best Local Similarity	63.9%;	Score 205.8; DB 6; Length 506;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;
FEATURES			
source			
BASE COUNT	131 a	84 c	125 g
ORIGIN			143 t
			23 others
Query Match	Best Local Similarity	97.0%;	Pred. No. 1.4e-49;
	Matches 229; Conservative	0;	Mismatches 5; Indels 2; Gaps 2;

```
Db 136 CCGTCAATCTGGTTATGAAACAACCTGAGACACCTTCTCCGATGGCCAGTATGTNC 195
OY 180 ccaggtatgttggtagccatcct-gacagtiagagccgatcactggaag 234
Db 196 CCAGGTTATGTTGTGACCCATNTCTGAGACAGTTAGAGCCGATATCACTGGGAG 251
```

Search completed: March 30, 2002, 12:59:18  
Job time: 21440 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:31:11 ; Search time 5438.31 Seconds  
(without alignments)  
9028.073 Million cell updates/sec

Title: US-09-867-034-3

Perfect score: 4569  
Sequence: 1 atgataataatcctcaacac.....tgcagctgcaatagcttag 4569

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estlda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estl:\*  
11: gb\_estc2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_iny:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1514.4	33.1	2933	12	AK008659 Mus muscu
2	1500.2	32.8	2915	12	AK007466 Mus muscu
3	808	17.7	2703	13	AF101616 AF101616
4	655	14.3	825	10	AU122701 AU122701
5	631	13.8	745	11	BG622206 BG622206
6	630.4	13.8	943	13	AO748956 HS-5574_A
7	629.6	13.8	810	13	AO629742 RPCI-11-4
8	620.2	13.6	768	13	AO740881 HS-2274_A
9	617.8	13.5	893	10	AU137259 AU137259
10	615.2	13.5	863	13	AO752204 HS-3571_B
11	611	13.4	653	10	AW361534 OV2-CT026
12	610	13.4	872	13	AO896432 HS-3134_A

13	608.6	13.3	860	13	AO747583	AO747583 HS-5533_A
14	606	13.3	796	13	AO636664	AO636664 RPCI-11-4
15	605.4	13.3	759	13	AO058259	AO058259 CIT-HSP-2
16	603.4	13.2	734	10	AI660957	AI660957 wF20408.X
17	603.2	13.2	756	13	AO041614	AO041614 CIT-HSP-2
18	595.6	13.0	897	13	AO746355	AO746355 HS-2278_A
19	592.8	13.0	820	13	AO746771	AO746771 HS-2276_A
20	589.4	12.9	793	13	AO542629	AO542629 RPCI-11-3
21	586.2	12.8	891	13	AO746363	AO746363 HS-2278_A
22	584.8	12.8	1393	12	AF116619	AF116619 Homo. Sap1
23	582.6	12.8	828	13	AO780863	AO780863 HS-3113_B
24	580	12.7	689	13	BI6435	BI6435 342B19.FPB
25	578.8	12.7	712	11	BG69667	BG69667 602681514
26	578.6	12.7	863	13	AO746062	AO746062 HS-2277_A
27	578.6	12.7	894	13	AO739350	AO739350 HS-5482_B
28	572.2	12.5	762	13	AO200047	AO200047 RPCI11-59
29	572.2	12.5	763	13	AO039481	AO039481 CIT-HSP-2
30	569	12.5	925	11	BG403000	BG403000 602418809
31	567.8	12.4	598	11	BE927204	BE927204 RCI-CT024
32	567	12.4	766	13	AO745976	AO745976 HS-2275_A
33	566.6	12.4	633	10	AW753451	AW753451 OV2-CT026
34	565	12.4	835	13	AO780914	AO780914 HS-3113_B
35	564.8	12.4	734	13	AO262912	AO262912 CIT-B1-E1
36	563.6	12.3	721	13	AO055831	AO055831 CIT-HSP-2
37	562	12.3	721	13	AO395312	AO395312 CIT-B1-E1
38	560.8	12.3	710	13	AO077037	AO077037 CIT-HSP-2
39	560.2	12.3	730	13	B59966	B59966 CIT-HSP-347
40	558.6	12.2	723	13	AO383358	AO383358 RPCI11-13
41	558.6	12.2	842	13	AO747087	AO747087 HS-5538_A
42	557.4	12.2	709	13	B92460	B92460 CIT-HSP-217
43	555.4	12.1	625	10	AW361532	AW361532 OV2-CT026
44	554.4	12.1	782	13	AO269601	AO269601 HS-2036
45	552.6	12.1	712	13	AO047421	AO047421 CIT-HSP-2

#### ALIGNMENTS

RESULT 1	AK008659	2933 bp	mRNA	HTC	05-JUL-2001
LOCUS	AK008659				
DEFINITION	Mus musculus adult male stomach cDNA, RIKEN full-length enriched library, clone:2210006G11, full insert sequence.				
ACCESSION	AK008659				
VERSION	AK008659.1	GI:12842987			
KEYWORDS	CAP trapper.				
SOURCE	Mus musculus (strain:C57BL/6J) adult male stomach cDNA to mRNA, clone:libR1REN full-length enriched mouse cDNA library clone:2210006G11.				
ORGANISM	Mus musculus				
REFERENCE	EunhyiYota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	1 (bases 1 to 2933)				
TITLE	Carninci, P. and Hayashizaki, Y.				
JOURNAL	High-efficiency full-length cDNA cloning				
MEDLINE	Methods in enzymology. 303, 19-44 (1999)				
PUBMED	99279253				
REFERENCE	10349636				
AUTHORS	2 (bases 1 to 2933)				
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
JOURNAL	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
MEDLINE	Genome research. 10 (10), 1617-1630 (2000)				
PUBMED	20499374				
REFERENCE	11042159				
AUTHORS	3 (bases 1 to 2933)				
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, Y., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, E., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Wataniki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,				





Db	968	AAGTCCGGAGACATGCTGGAACGATGATCGTCTTAACCCGATAAATTCAGGAACCGCGCTT	1021
Oy	2725	ttctctgcctgcagaaacagttgtgacgtggtgctcttggttggatgtgtacatlttbaacagttgct	2784
Db	1028	TTCTCGCTGCAGACTGTGGAGCGACGGGATCCTGGGTGGGATGGTGCCTTTACAGTCTCT	1087
Oy	2785	gcccatgtacaacaaatgaactacatagataaacaagttgcagttgcacgaaggacacactgcgc	2844
Db	1088	GCCATATGACAAACCGAACTCAAAACAGTTTAAACAGTGGTGCTGACAGAGATTCGTCTGATC	1147
Oy	2845	aaaagatbaacctgaagaagcttgaaggaggagtcacatgtgcagggtgcttgatcgcga	2904
Db	1148	AAGCACTTACCACAGTATCTGCAAGGGGAGCATCTATATGCTCTGGCCCTTGCGACACCA	1207
Oy	2905	tttactgatatgttgcacaacatltgcctgttlttccatgacacacagcaatltatvggagtg	2964
Db	1208	TTTTAC-----	1212
Oy	2965	cgacaagaanaatccaacatvtggcctctctgacctgcagcttagtgattagaagaanaat	3024
Db	1213	-----AGTATTAAGCAAGAGTAT	1231
Oy	3025	ccaactgatgatcttgaatlaattgtctgcctggaagatgtgggaagaacaaactataagttgg	3084
Db	1232	CCAACCTGATGATGTGAATTTGCTGCTGCACCGATGGGAGACCAACCCATTAGCGAGC	1291
Oy	3085	tgctttaacgagtgccaacaaagtgtgtccatcatccacacagtcgtcttggggccctct	3144
Db	1292	TGCTTTGACCGTGGTGAAGCGACAGCGGGGCCATATCCATACAGTGGCCCTGGGACCGGCT	1351
Oy	3145	gcagctcaagaactagagagagctgtgccaaatgacagagaggttacaagacatagtctta	3204
Db	1352	GCCCCTAAAGAGCTTGAGCGAGCTGTCCAAATATACAGAGGCCCTGCAGACATACCTCTCG	1411
Oy	3205	gataaaatltgaagaacaaatggccctcatgtatgtcttggggcccttcaatcaggaaatbga	3264
Db	1412	GATGATGTTCCAGAACAAATGGTCTTTGTTGATGCTTTTGCGACCTCTCCACAGAAATCCG	1471
Oy	3265	gctgtctctcagcgtctccatccacagcttgaagatgaagagatlaacctccagaacagccag	3324
Db	1472	GGCATCGCTCAGCACTCCATCCACGCTGAGAGCAGGGGAGTTAATCTCCAGAAATACCA	1531
Oy	3325	ttgatltgaatgagcacagtcatctgtgagacgaaccgttggaaagagacatttgttctatc	3384
Db	1532	TGGATGTGATGGCTCAGTACGATCGTGGACAGCTCGTGGCAAGACACCTTGTTCTTATTC	1591
Oy	3385	accctggacaatgacgcctccccaatccctcttggatcccaagtggagcagaagaagaagt	3444
Db	1592	ACCTGGACAAACGCACTCTCTTACAAATTTATTTGGGATCCAGCGAGGTGGAAACAAAT	1651
Oy	3445	gaccttgtagtggacaaacacacaaatbgtccctacatccaatcccaagcatttctga	3504
Db	1652	GGTTTATTACAGACAAACACCTAAGTGTGGCTTACTCTCAAGTCCGACGACGGCTAAG	1711
Oy	3505	gtttggacattggaaatacagttctgcaagaagaagctcaacaaacttgaacctgtagcttgcag	3564
Db	1712	GTTTGACTTTTGGAAATACACATTTCAAGGAGAGCTCACAGACTCTCACCTTACCTGTCC	1771
Oy	3565	tcccggtgcgtccaatgttaccctgtcctccaatatacagtgtaacttccaaaacgaacaagac	3624
Db	1772	TCCCGTCCAGCAAGTGTCTTACACTGCTCTCTATTACAGTACCCCGGTAGTGAATTAAGAC	1831
Oy	3625	accgaagaattccccacacctctgtgttagttatgaaatattcgccaagagagccctcccca	3684
Db	1832	ACAGGGAAATTTCCCGACGCTTGTACAGTGTATGACAGCAATTTGCCCAAGAGACCTCGCT	1891
Oy	3685	attctcagggccagttgtcacagcccttgatltgaatcagttgaatgtgaaaaaagattaccttg	3744
Db	1892	ATTCTCAGGGGCGACGTCACAGGCTTGATTTGAATCTGTGAATGAGAAAAACAGTAAACCTG	1951
Oy	3745	gaactacttgataatagagcaaggtgctgtatgtactaagaagatgacggtgtctactcaag	3804
Db	1952	GAATTACTGTGATTAACGGAGCAGGTGGCGATGCCACCAACAAATGAATGAGTGTCTACTCAAG	2011

QY	3805	tatttcacaactt	atgacagaaatggtatagatacaagtgataaagtctcgggctctggagga	3864
Db	2012	TTTTTTACAGCTTTT	TGAATGCAAAATGGTAGATACAGCCTTTAAATATATGGGCTCTGGAGGA	2071
QY	3865	gttaacgcagccacaacggaagatgataccccagagaaggtggaaacctgaactgaactgaactgac	3924	
Db	2072	GTCACTTTAGACACAGACAGAGACAGCACCTTCCGAGAACAGACCCATGTATCATATGATGGAC	2131	
QY	3925	tggaattgagaa	tgaataaacaatggaaatccacaagaacctgaataataagaatgat	3984
Db	2132	TGGATTTGAGGATGGGGAAGTAGAATGAAACCAACCACTGCTGTGAACCTGTT	-----AT	2185
QY	3985	gttcaacacaagcaagatggtgttlltagcagaacaatcctcggagagctcatattgtgtcttct	4044	
Db	2186	GTTCAGAACCAAGACAGAGCTGTGCTTAGAGAGAGACATCTTCAGGGGGATTCGTTTTGGGCGACCC	2245	
QY	4045	gatgtgc-----caaatgctcccatcaccgatctctctccaccctggccaaatcacagacctg	4101	
Db	2246	AATGTCGCCCGGACGACAGCTCCCATTTCTGTACTCTTTTCCACCTCTGTCAATACTGTGACTG	2305	
QY	4102	aagcgagaatcatcacggygggcagctcatcaatcaattgacttgacttgacagctctctgggagatg	4161	
Db	2306	AAGGCACATTCACCAAGGCGACAAACCTGTGATGTGACGTGACGGCTGCTGGGATGAG	2365	
QY	4162	tatgacctatgaaacagctctcaagaatatactatcttgaaataagataagatattcttgatctc	4221	
Db	2366	TACGACCCACGGGAGAGCTTCACACATCAATCATCCGAATGAGCACAGTATCGTTGATCTC	2425	
QY	4222	aagagcaagatccaatgaaatctctccaagttaatactactctctcatcccaagaagagacc	4281	
Db	2426	AGGACACACTTCACACACTCTCACTCTCCAGTAAACACTACCGGTCTTATCCCGAANAGAGGCC	2485	
QY	4282	aactctgaggaagtccttllgtllttaaaccagaaacattactctlltgaatatggcacagat	4341	
Db	2486	AGCTGTGAGAAATCTTTGATTTGAACTGGAGAGCAACACTTTTGGAAATGCGACAGAT	2545	
QY	4342	ctttcatctgctatctcagagctgtgtgataaagtgatcttgaaatccagaatatccaactt	4401	
Db	2546	ATCTTCATTTGCTATTCACAGGCTGTGGATTAATGATCCATTTGAAATCAAAAATCTGCAACATT	2605	
QY	4402	gcaacagatcatcttgattatctcccaacaagaactccgcagaagacacactagctctgata	4461	
Db	2606	GCAGCGGTCTGTGTGTATCTCCCGCTCAG-----GAGCGGCCATTTCCCGAAGAC	2656	
QY	4462	acgtctgctcctgtctcctaataatcatatacaacagacacatctccctggcatccaattta	4521	
Db	2657	TCAACTCCCTCTGCTCCATCATGAGATCAACAGCACATTCCTGGCATCCACGTGCTG	2716	
QY	4522	aaaattatgtgaagtgatgataaggaagactcagctgtcatatagacctag	4569	
Db	2717	AGATTAATGTGAAGATGGCTTAGGGGAATGAGGTGACACTAGAGTTTG	2764	
RESULT	2			
AK007466				
LOCUS	AK007466	2915 bp	mrna	HTC
DEFINITION	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length enriched library, clone:1810012P03, full insert sequence.			
ACCESSION	AK007466			
VERSION	AK007466.1	GI:12841032		
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to mRNA, clone:Jlib.RIKEN full-length enriched mouse cDNA library			
	Clone:1810012P03.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus			
REFERENCE	Carninci, P. and Hayashizaki, Y.			
AUTHORS	1 (bases 1 to 2915)			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Methods in enzymology. 303, 19-44 (1999)			
MEDLINE	99279253			

RESULT	2			
AK007466				
LOCUS	AK007466	2915 bp	RNA	HTC
DEFINITION	Mus musculus 10 day old male pancreas	cDNA, RIKEN full-length		05-JUL-2001
ACCESSION	AK007466			
VERSION	AK007466.1	GI:12841032		
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J)	10 day old male pancreas	cDNA to mRNA, clone:1lb.RIKEN full-length enriched mouse cDNA library	
			clone:1810012P03.	

ORGANISM	REFERENCE
Mus musculus	Carninci, P. and Hayashizaki, Y. (1999)
Euryarchaea: Methanococcus	High-efficiency full-length cDNA cloning methods in enzymology. 303, 19-44 (1999)
Chordata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sclerogammatia: Muridae: Murinae: Mus	99379253

PUBMED REFERENCE AUTHORS	10349636 2 (bases 1 to 2915) Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, H., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL MEDLINE PUBMED	20499374 11042159
REFERENCE AUTHORS	3 (bases 1 to 2915) Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, T., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujisake, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Katahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, T., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	Riken integrated sequence analysis (RISA) system—384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL MEDLINE PUBMED	Genome research. 10 (11), 1757-1771 (2000) 20530913 11076661
REFERENCE AUTHORS	4 (bases 1 to 2915) The Riken Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL REFERENCE AUTHORS	Nature 409, 685-690 (2001) 5 (bases 1 to 2915) Adachi, T., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A.,

TITLE	JOURNAL	REFERENCE
4. (bases 1 to 2915)		
THE RIKEN GENOME EXPLOSION		
FANTOM Consortium.		
Functional annotation of a full-length mouse cDNA collection		
Nature 409, 685-690 (2001)		
5. (bases 1 to 2915)		
AASHLI, J., AIZAWA, K., AKAHIRA, S., AKIMURA, T., ANO, H., ARAI, A.,		
ARAKAWA, T., CARNICOLI, P., FUKUDA, S., FUKUSHI, Y., FUNNO, M.,		
HANAGAKI, T., HARA, A., HAYATSU, N., HIMEMOTO, K., HIROKITA, T., HORI, F.		
IMOTOANI, K., ISHII, Y., ITOH, M., IZAWA, M., KATO, H., KAWAI, J.,		
KOJIMA, Y., KONDO, H., KUNDA, M., KOY, S., KURIHARA, C., MATSUYAMA, T.		
MIYAZAKI, A., NISHI, K., NOMURA, K., NUMAZAKI, R., OHNO, M., OKAZAKI, Y.		
OKADA, T., OGA, C., SAITO, H., SAITO, R., SAKAI, C., SAKAI, K., SANO, H.,		
SEKAKI, D., SHIBATA, K., SHIBATA, Y., SHINAGAWA, A., SHIRAKI, T.,		
SOGABE, T., SUZUKI, H., TAGAMI, M., TAGAWA, A., TAKAHASHI, F.,		
TANAKA, T., TEJIMA, Y., TOYA, T., YAMAMURA, T., YASUNISHI, A.,		
YOSHIDA, K., YOSHINO, M., MURAMATSU, M. and HAYASHIZAKI, Y.		
Direct Submission		
Submitted (10-JUL-2000)		

**TITLE** Direct Submission  
**JOURNAL** Submitted (10-JUN-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)  
**COMMENT** Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

**FEATURES**

**source**

Location/Qualifiers

1..2915

```
Location/Qualifiers
1. .2915
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/db_xref="MGD:1046342"
/db_xref="MGI:1893339"
/clone="1810012P03"
/sex="male"
/russus_type="pancreas"
/clone_id="Riken full-length enriched mouse cDNA library"
```

BASE COUNT	838 a	720 c	695 g	662 t	/dev stage="10 day old"
ORIGIN					
Query Match	32.8%;	Score 1500.2;	DB 12;	Length 2915;	
Best Local Similarity	74.9%;	Pred. No. 0;			
Matches 2013;	Conservative	0;	Mismatches 558;	Indels 118;	Gaps
QY	1888	aagacatggtgaccagcactcgtatcgtctgtgaagctacaggaagcatttat	1947		
DB	188	AAGGACATGTGAGATCTCAGGCTCTCCATACCTTTGGAAGCTACAGAAAGATTTC	247		
QY	1948	ttcgaatgttgcattttgtattccctgaacatggaacaaagctgtacatgtgga	2007		
DB	248	TTCAAAATGTGGCATTGTTGATTCGCGAAGGTGGAAGGCAAACTCTGAATATACGAGG	307		
QY	2008	ccaaactctgagacctcaaaaatgtctgaatgttctgttcgtgagctcactccagat	2067		
DB	308	CCAAACTTGAAACCTTCAAAAGCGTGAGTCTCTGTATCAACAACAGCCCTTAGGC	367		
QY	2068	aatgtagaaccttaccactgagcagatggtgcactgttgagagaaggtlgaagatccac	2127		
DB	368	AATGATGAGCCCTACACCGAACAATATAGGAGCATGTGGAGAAAGGGGATCGAGATTAC	427		
QY	2128	ctcactcctgatttcattctgaggaaaaaagatgctggaatgtgagccaaagtgagca	2187		
DB	428	CTGACTCTGACTCTTTAGCAGGAAAGAGCTGACTGACTATGGCCACAGACAGACC	487		
QY	2188	tttgtccatgagttgggtctcactacgaatgggagatcttgacgagctacataatgatag	2247		
DB	488	TTTGTCATGAGTGGGTCTACTTCGATGGGGAGTGTATGAAATACAAACAGCAGG	547		
QY	2248	aaattctactatcccaatggaagaataacaaagcgtgaatgttcagcgatctactggt	2307		
DB	548	AAGTTCTACTATTCCAAAGGAAAAACCCCAAGCAGTAGGTTCAGACCCATTCACGGT	607		
QY	2308	acaatgtagtaagaagatgtccagggagcagctgtctaaccaa--aagatgcacatc	2364		
DB	608	AAAAATCAAGTTCTGCGGTCCAGGGAGGCAGTGTATCATTACAGGAAAGTGTTATTC	667		
QY	2365	aataaagtaacagagactctatgaaaaagatgtgattgttctccaatcccgccagacg	2424		
DB	668	GACGAGATACGGGACGTATTAAGACAATGTGTATTGTACAGATCCACACCAAAAC	727		
QY	2425	gagagagctctctaagtgttgacaaactgtgatctatagtgtgaattcgtctcaaa	2484		
DB	728	GAGAGAGCTTCCATCAGTGTAAACCAAAATATCAATTTGTGTGAATTCGTACAA	787		
QY	2485	caaaacacacaagaagagctccaaacaagcaaatccaaatgtcaatctccgaagcaca	2544		
DB	788	AAAAATCACATCAAGAAGGCCCAATATGACCAAAACCAAGATGCATATCCGAAAGCAG	847		
QY	2545	tgggaagtgtccgtgtatcttgaggaacttaagaaaaaacctctcatgacaacagcca	2604		
DB	848	TGGGAATGTCATCAGGATCTGAGGACTTCAGCAAAACCACTCCATGCACGCCAGCA	907		
QY	2605	ccaaatcccaacttccatctgctgcagagattggaacaaagatgtgtgttgctgttac	2664		
DB	908	CTCGACCCACCTTCTCACCTACCTGCGCAAAATTTGGCAAAAGATTTGTGCTTAATTCTTAT	967		
QY	2665	aaatctggaagcactgacgactgtgtaaacgcgcctcaatcgcactgaaatcaagcaagcagct	2724		
DB	968	AAGTCCGGGAGCATGCTGAAACGATGATCGTCTTGACCGAATGAATCAGGCAAGCGGCTT	1027		
QY	2725	ttctcgtctcagacagtttgagctgggtccctcgtggttggaatggtgacatttgaagtgct	2784		
DB	1028	TTTCTGCTGACAGCTGTGGAGCGAGGATCTCTGGGTCCGAGTGTGACCTTGACAGTGTCT	1087		
QY	2785	ggccatgtacaaaatgaactcatcaagataaacagtggtgcaatgacagggaacactgcgc	2844		
DB	1088	GCCATGTACAAAGCAACTCAACAGTTTAAACAGTGGGGTCTGACAGAGATCTGCTGTATC	1147		

Db	2132	CTGGATTGAGGATGGCTGAAGTAAAGTGAACCCACACAGCTCTGTGAACACTGTT-----A	2185
QY	3984	TGTCACACAAAGCAAGTGTGTTCTCGAGCAAGAACCTCCTGGGAGGCTATTGTGCTC	4043
Db	2186	TGTTCAAGACAAACAGCTGTGCTTTCAGCAGACGATCTTTAGGGGGATCGTTGTGTGGCCAC	2245
QY	4044	TGATGTC--CAATGCTCCCATCCTGATCTCTCCCACTGGCCAAATCCAGCACT	4100
Db	2246	CAATGTCCCGCAGCAGCTCCCATTCCTGACCTCTTCCACCTGTCAATATCATGACCT	2305
QY	4101	GAAGGCGGAATTCACGGGGGCGAGCTTCATATCTGATCTGGACGCTCTGGGATGA	4160
Db	2306	GAAGCCAGCATCCAAAGGGCAGACCTGGTGAATCGAGCTGACGGCTCCGGGGATGA	2365
QY	4161	TTGACCAAGGACAGCAAGTATCATCTGCAAGTAAAGTAAAGTAAAGTAAAGTAAAGT	4220
Db	2366	CTAGACACACAGGAGAGCTTCCAACTACATCATCCGAATGACGACCATGATGTGATCT	2425
QY	4221	CAGAGCAAGTTCATGATGATCTCTCTCAAGTGAATGATGATGATGATGATGATGATG	4280
Db	2426	CAGGACACCATTCACACCTCACTCCAAAGGAACATCATCGGCTTATCCCAAGAGGC	2485
QY	4281	CAATCTCGAGGAGTCTTTGTTTAAACCAAGAAACATTAATCTTGAATATGACAGA	4340
Db	2486	CAGCTCGAGGAATCTTCAGTTTGAACCTGGGAGCAACCTTTGGAAATGGACAGA	2545
QY	4341	TCCTTCATGCTATCAGGCTGTGATAGTGCATGCAATCAAGAAATCCAAAT	4400
Db	2546	TATTTTATGCTATCCAGGCTGTGATAGTGCATGCAATCAAGAAATCCAAAT	2605
QY	4401	TGCAGAGTATCTGTTTCTCTCTCCCAAGACTCCGCAAGAGCAAGCTAGCTGATGA	4460
Db	2606	TGCACGGGTGTGTGTTCATCCCGCTCAG-----GAGCGGCCATTCGCCAAGA	2656
QY	4461	AAGCTGCT	4520
Db	2657	CTCAGCT	2716
QY	4521	AAAATATGTAAGTGTAGTAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	4586
Db	2717	GAGATATATGTGAGATGCTAGGAGGAATGACAGTACACTAGGTTTG	2765
RESULT	3		
LOCUS	AF101616		
DEFINITION	AF101616 Human Homo sapiens genomic clone plasmid 23.325P6, DNA		
ACCESSION	AF101616		
VERSION	AF101616.1		
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 2703)		
	Begler, G., O'Brian, K.C., Kim, Y.C., Schreiber, G. and Pitterle, D.M.		
	A 1.4-Mb high-resolution physical map and contig of chromosome		
	segment 11p15.5 and genes in the LOH1A metastasis suppressor		
	region		
JOURNAL	Genomics 55 (2), 164-175 (1999)		
MEDLINE	99134294		
COMMENT	Contact: Begler, G		
	Duke University Medical Center		
	Box 2610, MSRB, Room 117, Durham, NC 27710, USA		
	part of a 1.4 megabase contig including the LOH1A metastasis		
	suppressor region Bin B		
	Class: unknown.		
FEATURES	Location/Qualifiers		
source	1..2703		
	/organism="Homo sapiens"		

/db\_xref="taxon:9606"  
 /map="11p15.5"  
 /clone="pTWB233.32SP6"  
 /clone\_lib="Human"  
 BASE COUNT 1057 a 573 c 483 g 590 t  
 ORIGIN

Query Match 17.7%; Score 808; DB 13; Length 2703;  
 Best Local Similarity 89.7%; Pred. No. 1.3e-166;  
 Matches 924; Conservative 0; Mismatches 50; Indels 56; Gaps 3;

QY 152 aattctaccagaggtacaaagagaagaaatggtaccattccctctggaagtattacaata 211  
 Db 1 AATTCTACCAAGAGGTACAAAGAGAGAGTGTACCATTCCTTGAAACTATTCACAA 60  
 QY 212 taagaaaaagaggaatccctcccttaactcgtttatagggcaacatcatcctgatacaa 271  
 Db 61 TAGAAAAAGAGGAATCCTCCCTAATCTTTATGAGGCCAGCATCATCCTGTACCAA 120  
 QY 272 agcgggcaagagacacacacacaaagaagaatttagacaatatcttgatgaacatg 331  
 Db 121 AGCCTCCAGAGACACACAAACAAAGAGAGTGTACCAATATCCTGTGATGACATCG 180  
 QY 332 atgcaaaaatccctcaataaataactgcaaacgcgaatccagcgacatcaaaaagctta 391  
 Db 181 ATCCAAAAATCCTCAATAAATACTGGCAACATCGAATCCAGACACATCAAAAAAGCTTA 240  
 QY 392 tccacatgatacaagtgggtcatccctgagat----- 425  
 Db 241 TCCACCATGATCAAGTGGGCTTATCCTCCCTGCATGCAAGCGTGTTCACATACGAAAT 300  
 QY 426 -----aaccaagacaaacacacacatgatatctcaa 457  
 Db 301 CAATAAACATATCCAGCATATTAACAGAACCAAGACAAACACATATGATTATCTCAA 360  
 QY 458 tagatgcagaaaaagccttgacaaaattcaaacaccttcacgtctaaaaacctcaata 517  
 Db 361 TAGATCAGAAAAAGCCTTTGGACAAAATTCACAAATGCTTCATGCTTAAATCTCAATA 420  
 QY 518 aattagatatgtgtggacatactcaaaaataaagaagctctatgtgaagacacag 577  
 Db 421 AATTAGTATGATGGGACATATCTCAAAATTAAGAGCTATCTATGACAAACCCACAG 480  
 QY 578 ccaatatcatctgaatggygcaaaaacttgaagacatccctctgaaaacttgcacagac 637  
 Db 481 CCAATATCATACTGATGATGACAAAACATGGAACATTCCTTGAAACATGCGACAAGGC 540  
 QY 638 aggaatgcccctctcaacacactctatcaactagtcttggaagcttgcagagggcaa 697  
 Db 541 AGGGATGCCCTCTCTCACCACTCTCTAATCAACATAGTGTGTGTAAGTCTGGCCAGGCAA 600  
 QY 698 tttagcagagagaagaataaagaagtttcaatctagaagaagaagtaattgtccc 757  
 Db 601 TCAGGCAAGAGGAAGAAATAAGGCAATTCATTAGAAAAAGAG-AGTAAATTGTGCC 659  
 QY 758 tgtttgcaggtgacatgatgtgtatacttagaaaaaccccatctctcgaagcccaaatctcc 817  
 Db 660 TGTTCAGATGACATGATGTTATCTTAGAAAAACCCCATTTGTCGAGCCCAAAATCTCC 719  
 QY 818 ttaagtgtaagaacaacttcagaaaagttcagatatacaaatcaaatgtatacaaatcac 877  
 Db 720 TTAAAGCTGATTAAGCAACTTCAGCAAAAGTCTCAGATATATAAATCACTGTCCAAAAATTCAC 779  
 QY 878 aagcattccatatacaacaataacagagaagacccaatcatcgaatggaactcccat 937  
 Db 780 AAGTATTTCTATGACCAATTAACAGACAAACAGAGCCCAATCATGAGTGAATCTCCAT 839  
 QY 938 tcaaatctgcttcaaaagagaataaataacactagaatcccaacttaagaaggtatgtaag 997  
 Db 840 TCACAAATTCCTTCAAAAGAGATATAAATACCTAGGAATCCAACTTACAAGGAGTGAAG 899  
 QY 998 acccttcaagagaactcaaacacactgctcaatgaataaagaagatatacaaacaaat 1057

Db 900 ACCTCTCAAGAGAACTACAAACCACTGCTCATGATAAATAAGAGATACAAACAAAT 959  
 QY 1058 ggaagaacattccatgctcctcagtagaagaatacaatcatcgtgaagaatgcccactgc 1117  
 Db 960 GGAAGAACAATTCATGCTTATGGGTAGGAAGAATC-ATATCTGAAAAATGCTATCTGCG 1018  
 QY 1118 ccaagattat 1127  
 Db 1019 CCAAGGTAAT 1028

## RESULT 4

AU122701 825 bp mRNA EST 23-OCT-2000  
 LOCUS AU122701 MAMMAL Homo sapiens cDNA clone MAMMA1002934 5', mRNA  
 DEFINITION sequence.  
 ACCESSION AU122701  
 VERSION AU122701.1 GI:10947417  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS  
 1 (bases 1 to 825)  
 Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,  
 Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and  
 Isogai,T.

TITLE HRI human cDNA project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3952  
 Fax: 81-438-52-3952

Email: genomics@hri.co.jp  
 HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix  
 Research Institute; cDNA library construction: Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 FEATURES  
 Location/Qualifiers  
 source 1..825  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="MAMMA1002934"  
 /clone\_lib="MAMMA1"  
 /tissue\_type="mammary gland"  
 /note="Vector: pME18SFL3"

BASE COUNT 322 a 177 c 149 g 174 t 3 others  
 ORIGIN

Query Match 14.3%; Score 655; DB 10; Length 825;  
 Best Local Similarity 89.4%; Pred. No. 3.9e-133;  
 Matches 737; Conservative 0; Mismatches 33; Indels 54; Gaps 1;

QY 347 ataaatactgycgaacccgaatccagcagacatacaaaaagcttatccacatgatcaag 406  
 Db 1 ATAAATATCTGCGAAACCGAATCCAGACACATCAAAAAAGCTTATCCACCATGATCAAG 60  
 QY 407 tgggtccatccctggat----- 425  
 Db 61 TGGGCTTCATCCCTGGGATGCAAGGCTGTTCAATATACGAAATCAATTAATGTAATCC 120  
 QY 426 -----aaccaagaagaaaaaacacatgattatctcaatagatgcagaagaag 472  
 Db 121 AGCATATAACAGAGCCAAAGACAAAACACATGATTTATCTCAATATGATGCAAGAAAAAG 180  
 QY 473 ccttgcaaaaatccaacaccttcataaaccctcaataaataatagatatgatg 532  
 Db 181 CTTTGACAAAATTTCAACAACCTTCATGCTAATAAACTCTCAATTAATTAAGTATTGATG 240

QY	533	ggagcactctcaaaataaataagagctatctctgtgcaagagccagagccaattctactga	592
Db	241	GGACGATATTTCACAAATATAAGAGCTATCTTAGACAAACCCACAGCCAAATATCTACTGA	300
QY	593	atcggcacaacaactcggnaacatccctcttgaaaacctgcgcacaagaagatgcccctctc	652
Db	301	ATGGGCAAAACCTGGACACATCTCCCTTTGAAACATCGGCACAAAGAGGATGCCCTCTCT	360
QY	653	caccacccctatccaacataglttttggaagltcttgcgcaggccaattaggcagaagaag	712
Db	361	CACCCCTCTCTACTCAACTAGTGTGGAAAGTCTGGCCAGGGCATCAGCAGAGAGAAG	420
QY	713	aaataaaggttttcaatttagaanaaaggagatcaaatctgcccctgtttgcagtgaca	772
Db	421	AAATTAAGGGATATTCATTAATTAGAAAAAGGAAAGTCAAAATTTCCCTGTTCACAGACA	480
QY	773	tgatgtatcactagaaaaccccatctctcctagcccaaatctcottaagtgaataaga	832
Db	481	TGATGTGTTTATCTAGAAAAACCCATCATCTCAGCCCAAAATCTCCTTAGCTGTATAGCA	540
QY	833	acttcagaagaagctcagatatacaaaatcaatgttaacaaaatcacaagcatctcctataca	892
Db	541	ACTTCAGCAAAAGTCTCAGGATATCAAAATATCAATGTACAAAATATCAAGCATTTATACA	600
QY	893	ccaataacacagagaacacagagcccaatcatcgaatgaactcccatcacaatgtgtcctaa	952
Db	601	CCAAACAAAGACAAACAAAGAGCCCAAAATCAATGGGTGAACCTCCATTCACAATTCCTCAA	660
QY	953	agagataaataacttagaataccaacttacaagaagatgtgtaagaagactcttcagaaga	1012
Db	661	AGAGATAAATAACTTATAGGAATCCAACTTACAAAGGATGTAAAGAGACTCTTTCAGAGGA	720
QY	1013	actacaacaccactgtctcaatgaaataaagaagatatacaacaaatggaagaacattccat	1072
Db	721	ACTACAACACACTGCTCAAGGAATAATTAAGANGACNCAACAAATGGAAGAATTCAT	780
QY	1073	gtctctgtatgtaagaagaatcaatatctgtgtaaatggccatactg	1116
Db	781	GCTCATGGGTAGGAAGAATCAATATCGNAAATATGGGACTACTG	824
RESULT	5		
LOCUS	BG622206	745 bp	mRNA
DEFINITION	602646847F1 NIH_MGC_79	Homo sapiens	CDNA clone IMAGE:4768690 5'
ACCESSION	BG622206		mRNA sequence.
VERSION	BG622206.1	GI:13673577	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 745)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-remail.nih.gov		
	Tissue Procurement: CLONTECH Laboratories, Inc.		
	cDNA Library Preparation: CLONTECH Laboratories, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.llnl.gov		
	Plate: L10CM1632	row: m	column: 11
	High quality sequence stop: 734.		
FEATURES	Location/Qualifiers		
	1..745		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4768690"		

Query Match	Best Local Similarity	Score	DB	Length	745;
Matches	661; Conservative	0;	Pred.No. 7,1e-126;	Indels	0;
BASE COUNT	285 a	159 c	133 g	168 t	
ORIGIN					
444	catgatattcctaataagatgacagcaaaagccctttgacaataattcaacaccccttaagt	503			
1	CATGATTATCTCAATATGATGCGAAGAAAGCCCTTTGACAAATTCACACGCCCTTCATGCT	60			
504	aaaacccttcaataatagatattgatatltgtagacatatctcaaatataaagactacta	563			
61	AAAACCTCTCATTAGATTTAGTATGATGAGATGATATCTCAAAATTAATTAAGACTATT	120			
564	tggcaaaagccacagccaataatactatgtaattggcacaataactggaagcattcccttga	623			
121	TGAAAAACCCACAGCCAAATCATATGTAATGGGCCAAAACTGGAAGCAATCCCTTTGAA	180			
624	aacttgacacagaagaaggaatggccctctccacacactccattcaacatatggtttggaat	683			
181	AACTGGCACAAACAGAGGATGCCCTCTCTCACCACTCTTAATTCACATAGTGTGGAAAT	240			
684	tctggccaggaagaattaggaagcagaagaagaaataaagggttttcaattagaaagaaga	743			
241	TCTGGCCAGGCAATTAGCGCAGAGAGAAAGAACTAAAGGGTATTCAATTAGAAAAAGGA	300			
744	agtcaaatgtccctgtttgcaagtgatgattgtatactcctagaanaaccccaattcttc	803			
301	AGTCAAAATGTCCCTGTTTGGACAGCACATGATTTGTAATCTTGAAGAAACCCATCGTTC	360			
804	agcccaaaaatctcctaagctgataagaacacttcagaagaacttcagaatgataaatacaa	863			
361	AGCCCAAAATCTCTTAACCTGATAGCAACTTCACCAAGTCTCGATGATACAAATTCAA	420			
864	tgtacaaaatcacagaactctctatcacaaataacagagaagaagagccaatcat	923			
421	TGTGCAAAATTCACAAAGACTCTTATTATACCAATATATACAAACGAGCAACCAATCAT	480			
924	gaatgaactccattacaacttgcttcaagaagaataaataactcctagaatccaacttac	983			
481	GAGTGAACCTCCATTCACATATGATTCAGAGAAATTAATTAATCTGGAATCCCAACTAC	540			
984	aaggagatggaaggaccttccaaggagagacttacaacacacgctcaatgaataaaga	1043			
541	AAAGGACGTGAAGGACCTTTCAAGGAGAACTPACAAACACCTGCTCAATGAATGAAAAA	600			
1044	gagatacaacaatggaagaacaattccatgcttcattgataaggaagaatcaatatctgta	1103			
601	GGACACAAATGGAAGAAACATTCCTCAATGCTCATGAGATAGGAGAAATCAATATCTGAA	660			
1104	aatggccatactgcccagaagattatgctagatataaaggatlatcattagg	1154			
661	AATGGCCATACTGCCAAGTAAATTTACAGATTCAATGCCATCCATCAAG	711			

DEFINITION HS\_5574.A2.F10.T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=1150 Col=20 Row=K, DNA sequence.

ACCESSION AQ748956

VERSION AQ748956.1 GI:5536114

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 943)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 9380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering\_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu Plate: 1150 Row: K Column: 20 Seq primer: T7 Class: BAC ends High quality sequence stop: 943. Location/Qualifiers

FEATURES

source 1..943

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Plate=1150 Col=20 Row=K"

/clone\_lib="RPCI-11 Human Male BAC Library"

/sex="male"

/note="Vector: pBACe3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites" 4 others

BASE COUNT 345 a 220 c 159 g 215 t

ORIGIN

Query Match 13.8%; Score 630.4; DB 13; Length 943; Best Local Similarity 88.2%; Pred. No. 9.8e-128; Matches 763; Conservative 0; Mismatches 44; Indels 58; Gaps 5;

QY 20 cataactctcccaagcctaaccagaagaagtgaattctctgaatagaccataacag 79

DB 23 CATACACCCCTCCCAAGCTAACACAGAAAGTTGAATCTCTAAAGACTAATAACAG 82

QY 80 gctctgatcttgagcaataatcaagaagcttcaaccacaaagaagccagaccagatg 139

DB 83 GCTCTGATATTGGCAATATCAATAGCTTACCAACCAAAAAAGTCCAGACCAATG 142

QY 140 gattcacagctgaattctaccagaggtacagaagaagactggtacacattccctctgaag 199

DB 143 GATTTCAGCGCAATCTTACCAGAGGTACAAGAGGAGCTGTACATTCCTCTGAATG 202

QY 200 tatatacaataagaagaagaaggaactctccttaactcgtttatagagccaactca 259

DB 203 TATTCCAACTATAGAAAAGAGGAACTCCCTTAATCAATTATAGAGCCGACATCA 262

QY 260 tctgtatatacaagccggagagagacacacaaagaaagaagatttagaccatactct 319

DB 263 TTCTGATACCAAGCTTGGAGAGACACAAACAAAAAGAAATTTAGACCAATATCCC 322

QY 320 tgaatgaacattgtagcaaaaacccctcaataaaactctgcaaacccgaatccagcagaca 379

DB 323 TGATGAACATGATGATGCAAAAAATCCTCAATAAATAGCGCAAAACCGATCCAGACACA 382

QY 380 tcaaaaagcttaccacccatgacatgaatggctctcctccctgggat----- 425

DB 383 TCAAAAAGCTTATTCACCATGATCAAGTGCGCTTTATTCCTGGGATGCCAAGCTGTTCA 442

QY 426 -----aaccaaagacaaaacaca 445

DB 443 ACATACCAAAATCAATTAATGTAATCCAGCATATAACACAGACCAAAAGACAAACACACA 502

QY 446 tgaatctcaataagatagtcgaagaagccttgacaaaattcaacaac-cccttaagtcta 504

DB 503 TGATTTATCTCAATAGATGACGAAAGGCTTTGACAAAATTCACACGCCCTTCATGCTTA 562

QY 505 aaaaacctcaataatagatattgtagggacatctcgaataataaagatctat 564

DB 563 AAAACTCTCAATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 622

QY 565 ggcacaaagccacagcccaatatactactgaaatgggcaaaaactgggaagcatcccttgaag 624

DB 623 GACAAACCCACAGCCATATCATTAATGAGGCAAAACCTGGAGCATCCCTTTGAAA 682

QY 625 actggcacaagaagagatgcctctctcaccacccctatcaacatagtttg-aaagt 683

DB 683 ACTGGCACAAGACAGGATGCTCCTCTCACACTCTTATCAACATAGTGTGATATGT 742

QY 684 tctggccagagcaattaggagagaggaagaagaatgaaggttttcaattagaaagaaga 743

DB 743 TCTGGCCAGGCAATTAATGAGAGGAGAGAAATNAGGCAATTAATGAGAAAGAGGA 802

QY 744 agtcaaatgtccctgttttcaggtgacatgattgtatatactagaaaacccatctctc 803

DB 803 AGTCAATTTGCTCCCTGTTTCAGAT-ACATTAATTTGATATTATTAATCCCATGCTC 861

QY 804 agcccaaatctccttaagctgata 828

DB 862 AGGCC-AAATCTCTTAAGCTGATA 885

RESULT 7

AQ629742 810 bp DNA 17-JUN-1999

LOCUS RPCI-11-467B15.TV RPCI-11 Homo sapiens genomic clone RPCI-11-467B15

DEFINITION , DNA sequence.

ACCESSION AQ629742

VERSION AQ629742.1 GI:5092377

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 810)

AUTHORS Zhao,S., Adams,M.D., Niernan,W., Malek,J., de Jong,P. and Venter,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL Unpublished (1997)

COMMENT Other\_GSSs: RPCI-11-467B15.TV

Contact: Shaying Zhao, William Niernan, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbeetlgr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page:



Query Match	13.5%;	Score 617.8;	DB 10;	Length 893;
Best Local Similarity	87.1%;	Pred. No. 5.6e-125;		
Matches 765;	Conservative	0;	Mismatches 54;	Indels 59;
				Gaps 6;

OY	1	atgataaattccctcaacaacataaacctcccaagacttaacaggagaagattgtaact	60
Db	17	ATGATATACATTTCCTCGACACATACACTCTCCCAAGACTTAACCAAGAGAAGATTGAATCT	76
OY	61	ctgaatagaccataacagagctctgatatgtgycgaalaalcaagagcttaaccaccaa	120
Db	77	CTGAATAGACCATAATMACAGAGCTGGAATTTGTGGCAATTAATCTTTTATCCACCCAA	136
OY	121	aagagtcagagcaagagctgtaatacagagctgaaattctaccagagtgtaacagaggaactg	180
Db	137	AAGAGTCAGAGCAAGATGGATTGTACAGCTGAATTTCTACAGAGGTTCAGAGAGGAACTG	196
OY	181	gtaccattccctctgaaagtattacaatcatagaaaaagaggaactccctccctaactcg	240
Db	197	GTACCAATTCCTTGTGAAGACTATTCATCATATGAAAAAGAGGAATCTCCCTAACTCA	256
OY	241	tttataagagccaacatcalccctgatatccaaaagccgycagagacataaccacaaaaagag	300
Db	257	TTTTATAGGCGACCATCATCTGTATACCAAGCGGGGAGAGCACACACCAAAAAAGAG	316
OY	301	aattttagaccaatctcttgatgaacattgatacgaataaactccatataactatctgca	360
Db	317	AATTTTAGACCAATATCTCTTGATGATACATTGATGCAAAAATTCCTCATTAATAATAGTSCA	376
OY	361	aacccaatccagagcacatcatcaaaaaagcttatccacatgataagtgggcttcatccct	420
Db	377	AACCGAATCCAGCGACATCAAAAAACCTTATCCACCATGATCAAGTGGGCTTCATCCCT	436
OY	421	gggaat-----a	426
Db	437	GGGATGCAAGCGCTGTTTCATATATCGCAAGTCAATAATGTAATCCAGCATATAAACAGA	456
OY	427	accaagaacaaaaacacatgattatctcaatagatgcagaaaaagccttgacaaat	486
Db	497	GCCAAAGCANAAMACCATGATATCTCAATAGATGCAGAAAAAGCCTTTGACAAAATT	556
OY	487	caacaacccctcatgctcaaaaaacccctaaataaattagatatgtatgtagaacattctcaaa	546
Db	557	CAACACCCCTTCATGCTPAAAAACTCTCAATPAAATTAAGTATGTGAGGAGTATTTTCAAA	616
OY	547	ataataagagctatctatgycagagccacagccaatactcaactgtaatggyccaanaact-	605
Db	617	ATATATAGAGCTATCTMTGACAAACCCAGNCAATATCATAGATGAGTGGCAAAAACTG	676
OY	606	ggaagcattccctcttgaaactgycacagaaga- ggaatgcctctctcaacactccat	664
Db	677	GGAGAGCTCCCTTTGAAAATGTGGCACAGACAGGGGATCCCTCTCTCAACCTCTCAT	736
OY	665	tcaacatagatttggaagcttcgcagaggaactagagc- aggaagaagaaatcaaaaggt	723
Db	737	TCAACATATGTTTGGAGTTCTGGGCAAGGAATCAGGCAAGGAGGAAGGAATAAAGGT	796
OY	724	tttcaattagaaaaaggaagtaaaatttccctgtttgcagtgagcat- gatgtata	782
Db	797	ATTCATTTGGGAAAAAGAGGAGTC- AATTGGCCCTGTTTGACAGCATGATGGTTA	855
OY	783	cttagaaaaaccccatctctcagcccaaatctcccta	820
Db	856	NCTAGAAAACTCCATCATATNTAAGCCAAATNNCCCTTA	893

RESULT	10	
LOCUS	A0752204	
DEFINITION	A0752204 863 bp DNA	19-JUL-1999
ACCESSION	HS-5571.B2.H12.T7A RPeI-11 Human Male BAC Library	Homo sapiens
VERSION	A0752204	genomic clone Plate=1147 Col=24 Row=P, DNA sequence.
KEYWORDS	A0752204.1	GI:5539362
	GSS.	



QY	457	atagctgcagaagaagcctttgacaaatctcaaacaccttcatgtcctaaccctc	516
Db	374	ATAGATGAGAGAAAAAGCCCTTTGACAAATTCAAACACCCCTTCATGCTCTCAAT	433
QY	517	aattagatattgtaggggacatctccaataatcaagaagctatctatgagcaagcna	576
Db	434	AAATTAGTATTGATGGGACATATTTTCAAAATTAATGAGGCTATCTATGACAAACCACA	493
QY	577	gccaatcatctactgtaatgtaggcaaaaactgtagaatcttcctttgaaaactgacaa	636
Db	494	GCCATATTCATCTACTAAACGGGCAAAAACGTAAGCAAGATTCCTTTGAAACTGGACAA	553
QY	637	caggatagcctctctcaccactcctatccaacatagtttggagttctggccaggga	696
Db	554	CAGGATAGCCCTCTCTCACACCTCTCATTCACATAGTGTGGAAGTCTGGCCAGGCA	613
QY	697	attagcagagaaagaataaagggtttcaattagagaagaggaatcaaatgttc	756
Db	614	ATTAGCGAGAGAGGGAATTA--GGGTATTCATTAGGGAAGAGGAATCAATATGTCC	671
QY	757	ctgtttgcagtgagcatgattglatacctagaaaacccctctctcagcccaaatctc	816
Db	672	CTGTTTGCAGAGACATGATGTATATCTAGAAANCCCATGCTCACCCCAAAATCTT	731
QY	817	cttaagctgataagcaacttcagcaaatgtcctaagatatacaaatcaatgataaata	876
Db	732	CTTAAGCTGATATAGCAACCTTTCAGCAAGCTCTCAGATCCCAATCATATACAAAATTA	791
QY	877	caagatctcctatcacccatcacagagaagaagagcccaatctgaatgaactccca	936
Db	792	CAAGCTTTTATTTCNCCCCACACCCAGCACAGAGAGCCCAATTTCTGAGTGACTTCCA	851
QY	937	ttcacatctgtc 948	
Db	852	TTTCACATGTGT 863	
RESULT 11			
AM361534/C			
LOCUS	AM361534	653 bp	mRNA
DEFINITION	QV2-CT0261-261099-011-f01 CT0261		EST
ACCESSION	AM361534		04-FEB-2000
VERSION	AM361534.1	GI:6866288	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 653)		
JOURNAL	HCBP <a href="http://www.judwig.org.br/ORESTES">http://www.judwig.org.br/ORESTES</a> .		
COMMENT	The FAPESP/LICR Human Cancer Genome Project Unpublished (1999) Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: <a href="mailto:asimpson@judwig.org.br">asimpson@judwig.org.br</a> This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.judwig.org.br/scripts/gethtml2.pl?pl1-QV2&amp;t2-QV2-CT0261">http://www.judwig.org.br/scripts/gethtml2.pl?pl1-QV2&amp;t2-QV2-CT0261</a> ) 261099-011-f01&t3=1999-10-26&t4=1) Seq primer: puc 18 forward High quality sequence start: 37 High quality sequence stop: 652. Location/Qualifiers 1..653 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="CT0261" /dev_stage="Adult"		
FEATURES	source		

/note="Organ: colon; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 /16 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 140 a 156 c 166 g 191 t  
ORIGIN

Query Match 13.4%; Score 611; DB 10; Length 653;  
Best Local Similarity 98.9%; Pred. No. 1.7e-123;  
Matches 656; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

QY 3317 acagcagatgagatgacagatgacagatgacagacacgagagagagacattgt 3376  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 644 ACAGCCAGTGA -GAATGACAGATGATCTGGACAGACACCGTGGAAAGACATTGT 586  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3377 tctatcaccttgacaatgacagcctcccaatccttcttgagatccagtgacaga 3436  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 585 TTCTTGACCTGGACAGACAGCTCCCAATCTTCTGGAGATCCAGTGACAGA 526  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3437 agcaaggttgcttctgtagtgacacacacacacacacacacacacacacacac 3496  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 525 AGCAAGGTGCTTTGATGACAAAAACAAAAAGGCTACCTCAATCCAGGCA 466  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3497 ttgctaagcttgacatttgacacacacacacacacacacacacacacacac 3556  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 465 TTCTTAAGTGGACATTGGAATACAGTCTGCAGACACCTCAGAACCTTGACCTGA 406  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3557 ctgtcacgtccgtgctgccaatgctacccctgctcccaatgacatgacttccaaaaga 3616  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 405 CTGTCAAGTCCGCGTCCCAATGCTACCTGCTCCCAATTAAGTACCTCCAAAAAGA 346  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3617 acagaagac 3676  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 345 ACAGAGACACACAAATTCACAGCCCTGTGAGTTTGCAAAATTCGCGCAAGAG 286  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3677 ccccccacatccacagagcagatgacacacacacacacacacacacacacacac 3736  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 285 CCGCCCAATTCACAGGCGCAGTGTACACACCTGATTGATTAAGTGAAGAAAAAG 226  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3737 ttaaccttgac 3796  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 225 TTACCTTGGAACTACGATTAATGAGACAGTGTCTGATCTACTAAGAGATGACGGTGTCT 166  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3797 actcaagatattcaacactatgacacacacacacacacacacacacacacac 3856  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 165 ACTCAAGATTTTCACAACTTATGACACGAATGTAGATACAGTGTAAAGTCCGGGCTC 106  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3857 tgggaagagttaacagcagacagacagagatgatacccccagagtgagacacatgaca 3916  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 105 TGGGAGAGATTACGACGACGACGAGAGATGATACCCGACGACAGTGGACACTGTACA 46  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 3917 taactgg-ctgagatgagaatgataaatacaatgaaatcaac 3958  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 45 TACCTGGACTGATGAGATGATGAATACATGCAATCACTCTC 3

RESULT 12  
AO896432/c 872 bp DNA GSS 10-NOV-1999  
LOCUS HS.3134.A1.A11.T7C.CIT Approved Human Genomic Sperm Library D Homo  
DEFINITION sapiens genomic clone Plate=3134 Col=21 Row=A, DNA sequence.  
ACCESSION AO896432  
VERSION AO896432.1 GI:6352622  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 872)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
TITLE Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589  
COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones may be purchased from Research Genetics (info@resgen.com).  
BAC end Web Server: http://www.htsc.washington.edu  
Plate: 3134 row: A column: 21  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 872.  
Location/Qualifiers  
1. 872

FEATURES  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=3134 Col=21 Row=A"  
/clone\_1b="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in  
E-Coli DH10B"

BASE COUNT 197 a 155 c 172 g 342 t 6 others  
ORIGIN

Query Match 13.4%; Score 610; DB 13; Length 872;  
Best Local Similarity 93.5%; Pred. No. 2.9e-123;  
Matches 657; Conservative 0; Mismatches 43; Indels 3; Gaps 2;

QY 424 ataccac 483  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 710 AGAACCAAAAGACAAAAACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 651  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 484 atccac 543  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 650 ATCCACAAACGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 591  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 544 aaataataagagatct 603  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 590 CAATATATAGACTATCTATGACAAACCCACAGCCCAATATCATGATGATGATGATGAT 531  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 604 ctggaagcattcccttgaaactgac 663  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 530 CTGGAAGAAATTCCTTTGAAACTGGCACAAGACAGGATGCCCTCTCACCACCTCTTA 471  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 664 ttaacataatcttggaagctctgagcagagcaatttaggaagagaagaaataaaggt 723  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 470 TTCAACATTAAGTGTGGAAGTGTGGCCAGGCAATCAGCAGAGAGAGAGAAATTAAGAGGT 411  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 724 ttccaattagaagaagagagac 783  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 410 ATTCAATTAGAAAAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 353  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 784 ctagaacaccccaattctcagcccaac 843  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 352 CTGAAACACCCCAATGCTGACGCCCAAAATCTCTTCAGCTGATTAAGCACTTCAGCAAA 293  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 844 gttcagagatacaaaatcaatgac 902  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 292 GTCTCAGGATACAAAATCAATGTGCAAAAATCAAGCATCTTTATACACCAATATATAA 233  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
QY 903 agaaacagagac 962  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 232 ACAAAAAGAGCCCAATCTATGAGTACTCCATTACATATTGCTTCAAAAGAGATATAA 173

Qy 153 attctaccagaagtaaca-ggaaggaactgtaacattccctctgaagaagtattacaatcaa 211  
|||||  
Db 19 attctaccagaagttcaaggaggaactgtaacattccctctgaagaagtattccaatcaa 78  
|||||

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 796)  
REFERENCE  
AUTHORS Zhu,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter  
,J.C.

TITLE Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Other\_GSSs: RPCI-11-479A11.TU  
 Contact: Shaying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbe@tigr.org  
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: T7  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1. 796  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7683562"  
 /db\_xref="taxon:9606"  
 /clone="RPCI-11-479A11"  
 /clone\_lib="RPCI-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: PBACE3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCI11 Human Male BAC Library"

BASE COUNT 302 a 178 c 131 g 185 t

ORIGIN

Query Match 13.3%; Score 606; DB 13; Length 796;  
 Best Local Similarity 87.6%; Pred. No. 2.1e-122;  
 Matches 697; Conservative 0; Mismatches 45; Indels 54; Gaps 1;

QY 153 attctaccagaaggtacaaaggaaggaactggtacacatccctctgtaaaatatacaatcaat 212  
 1 ATTTCTCCAGAGGTACAAAGAGAGAGCTGTACATTCCTTCTGAAACTATATCCAACTAT 60  
 Db 213 agaaaagaaggaacatctcccttaactggtttatagaggaacatactccttgataccaa 272  
 61 AGAAAAATAGAAATCTCCCTAATCATTTTATGAGGCTAGCATCATCCTGATACCAAA 120  
 QY 273 gccggcagagacacaaacaaagaagaaatttagaacaatatcttgaatgaacattga 332  
 121 GCCTGCGAGAGACACAAACAAAAGAGATTTTAGACCAATATCCTGTGAAACATCAA 180  
 Db 333 tgcaaaaatcctcaataaatactgagcaaacgaaatccagcaagcacatcaaaaagcttat 392  
 181 TGCAAAAATCTCAATTAATTAATCTGGCAACCAATTCACACATATCAAAAAAGCTTAT 240  
 QY 393 cccacatgatcaagtgggtctcatccttgat----- 425  
 241 CCACCATGATCAAGTGGGTTTCATCCCTGGATGCAAGGCTGTTCAACTTATGCAATC 300  
 Db 426 -----aaccaagaacaaaaccacatgaatltatcgaat 458  
 301 AATTAACGTATATCCAGTGTATTAACAGAACCAACGACAAAACACATGATTTATCTCAAT 360  
 QY 459 agatgcagaanaagccttgcacaanaatlcacacacccctcatcgttaaaaacccctcaataa 518  
 361 AGATGCGAGAAAAGCCTTGACAAAATTCACAGCCCTCATCTGTAATAAACTCTCATATA 420  
 Db 519 attagatatgtatggtacacatatctcaaaataataagagctatctatggaagaagccacagc 578  
 421 ATTAGTATTTGATGAGACACATATCTCAAAAATTAATAGAGTATCTACGACAAACCCACAGC 480  
 QY 579 caatcacatcgaatattgggcaaaaactggaagatccctcttgaanaactggaacagaca 638  
 481 CAATATCATCTACTGAATGGGCAAAAACCTGGAAGCATTCCTTTGAAAACTGGCACAAAGACA 540

QY 639 gggatgccctctctcaccactcctattcaacatagtttggaaagttcgtgcccagggcaat 698  
 541 GGGATGCCCTCTCTCACCACCTCCATTCACACATAGTGTGGAGTTCTGGCCAGGGCAAT 600  
 QY 699 taaggcagagaaggaataaagggttccaattagaagaagaaggaatgaattgtccct 758  
 601 TAGCAGAGAGAGAAATTAAGGGTATTCAATTAAGGAAAGATGAAGTCAAAATGTTCCT 660  
 QY 759 gtttcgagtgacatgattgtataccatgaagaaccccatctcgaagcccaaatctcct 818  
 661 GTTTGCGATGACATGATTTGATATCTAGAAAAGCCCATATCTCAGCCCAAAATCTCCT 720  
 QY 819 taagtgataagcaacttcagcaagtlctcaggtatatacaaaaatcaatgatacaatcaca 878  
 721 TAAGCTATTAAGCAACTTCAGCAAGTCTCATGATACAAATGATGTGCAAAACTCACA 780  
 Db 879 agcaatcctataacc 894  
 781 AGCATTTCTTATACACC 796

RESULT 15  
 A0058259 759 bp DNA GSS 30-JUL-1998  
 LOCUS C1T-HSP-2345F5.TF C1T-HSP Homo sapiens genomic clone 2345F5, DNA  
 DEFINITION sequence.  
 ACCESSION A0058259.1 GI:3354785  
 VERSION A0058259.1  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 759)  
 Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,  
 Golden,K., Berry,K., Granger,D., Suh,E., Wilde,C., Shizuya,H.,  
 Simon,M. and Venter,J.C.  
 Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)  
 JOURNAL Unpublished (1998)  
 COMMENT Other\_GSSs: C1T-HSP-2345F5.TF  
 Contact: Mark Adams  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: mdadams@tigr.org  
 Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:  
 http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: M13 Reverse  
 Class: BAC ends.

FEATURES  
 Location/Qualifiers  
 1. 759  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="2345F5"  
 /clone\_lib="C1T-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2: HindIII"

BASE COUNT 296 a 158 c 136 g 169 t

ORIGIN

Query Match 13.3%; Score 605.4; DB 13; Length 759;  
 Best Local Similarity 95.6%; Pred. No. 2.9e-122;  
 Matches 634; Conservative 0; Mismatches 26; Indels 3; Gaps 1;

QY 424 ataaccaagacaaaaccacatgatattctaatagatgcagaagaagccttgcacaa 483

```
Db 100 AGAACCAAGACAAAACACATGATATCTCAATAGATGCAGAAAGCCTTGACAAA 159
QY 484 attcaaacacccctcatgcttaaaaaacccctaaataatagatatgtgagacatatc 543
Db 160 ATTCAACAGCCCTTATGCTAATAAACTCTCAATAATTAGTATGATGGACGTATCTC 219
QY 544 aaaaataaagagctatctatgcaagaagcacaagccaatcatatctaatgaggcaaaa 603
Db 220 AAAATPAATPAAGAGCTATCTATGACAAACCAAGCCATATCATATTGAATGGCAAAA 279
QY 604 ctggaagcatccctctgaaaaactgycacaagaaggaatgcccctctcaccactcta 663
Db 280 CTGGAGAGCATTCCTTTGAAAAACGGGCACAAGACAGGGATGCCCTCTCAACACTCCTA 339
QY 664 ttcaacatagttttggaaagttctgcccaggccaattagcaggagaagaagaataaggt 723
Db 340 TTCAACATATGTTGGAGTGTGCGCCAGGCAATTAGGCAAGAGAAATPAAGGCT 399
QY 724 ttcaatlaggaagaaaggaagtcacaaattgtccctgttgcaggtgacatgatatc 783
Db 400 ATTCAATTAGGAAAAAGAGAGAGTCAATTGTCCGTGTTGCAGATGACATGATTGTATAT 459
QY 784 ctagaaaaacccatctctcagcccaaaatctccttaagctgataagaacttcagcaa 843
Db 460 TTAGAAAAACCCATCGTCTCAGCCCAAAATCTC--AAGCTGATCAGCACTTCAGCAA 516
QY 844 gtccaagatcaaaaatcaatgtaacaaaatcacaagaatctcctataccaataaaga 903
Db 517 GTCTCAGGATACAAAATCAATGTAACAAAATCACAAAGCATTCTTATACCAATPAACAGA 576
QY 904 gaacaagaagccaatcatgtaatgaactcccatcacaattgcttcaagaagaataaa 963
Db 577 CAACAGAGAGCCCAATCATGAGTCAATTCCCATTCACAAATTGCTTCAAGAGATTAATA 636
QY 964 tacctaaggaatccaacttacaagggaatgtaaggaactcttcaaggagaactacaa 1023
Db 637 TACCTAGGAATCCAACCTACAGGATGTGAAGGACCTTTAAGGAGAACTACAAACCA 696
QY 1024 ctgctcaatgaataaagaagatacaacaataatggaagaacatctcagctcatgata 1083
Db 697 CTGCTTAATGAATPAACAGGATACAAACAATGGAAGAAACATTCCATGCTCATGGGTA 756
QY 1084 gga 1086
Db 757 GGA 759
```

Search completed: March 30, 2002, 09:32:04  
Job time: 13981 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:05:03 ; Search time 519.52 Seconds  
(without alignments)  
7539.881 Million cell updates/sec

Title: US-09-867-034-3

Perfect score: 4569  
Sequence: 1 atgagtaaatctctcaacac.....tcgacgtctgcatagcctag 4569

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: N\_Geneseq\_1101.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2492.2	54.5	3311	AA129502	C902P determined c
2	2491.2	54.5	2854	AAH34879	Human colon cancer
3	2491.2	54.5	2854	AAH31787	Human secreted pro
4	2489.2	54.5	3111	AAZ09840	Human membrane spa
5	2485.6	54.4	2825	AAH46124	Human CICAL CDNA,
6	2477	54.2	2867	AAH33285	Human colon cancer
7	2473.6	54.1	2745	AAH31927	Human ICAC-1 nucl
8	2469	54.0	2742	AAH46102	Human CICAL coding
9	2467.6	54.0	3109	AAH35019	Human colon cancer
10	1517.6	33.2	2739	AAH46101	Human Gob-5 coding
11	1517.6	33.2	2843	AAH46120	Mouse Gob-5 CDNA,

12	1512.8	33.1	2931	20	AAH81925	Murine ICAC-1 nuc
13	1129.8	24.7	3265	21	AAH26095	Membrane-bound pro
14	1129.8	24.7	3265	22	AAH20922	Human PRO1124 CDNA
15	1129.8	24.7	3265	22	AAH4241	Human PRO1124 (UNC)
16	1015.2	22.2	4709	21	AAH69112	Human secreted pro
17	1012.4	22.2	5065	21	AAH69111	Human secreted pro
18	1012.4	22.2	5065	22	AAH69112	Human secreted pro
19	995.8	21.8	32042	20	AAH209252	Human secreted pro
20	995.8	21.8	32042	20	AAH209252	Human secreted pro
21	992.6	21.7	49999	20	AAH30011	Human CARD-4 gene.
22	984.6	21.5	32199	22	AAH23901	Human LOBO homolo
23	981.4	21.5	50000	22	AAH57673	Human colorectal c
24	981.4	21.5	81145	22	AAH54867	Nucleotide sequenc
25	970.2	21.2	40328	21	AAH54868	Genomic nucleotide
26	969.4	21.2	14460	21	AAH23815	Human DAZ genomic
27	964.6	21.1	7207	22	AAH33815	Olfactory receptor
28	957.6	21.0	26190	22	AAH33404	Human kidney relat
29	957.4	21.0	18596	22	AAH31279	Human kidney relat
30	957.4	21.0	18596	22	AAH31215	Thymidylate syntha
31	940.6	20.6	23855	22	AAH3761	Human thymidylate
32	939.2	20.6	11520	22	AAH88291	Human kidney relat
33	937	20.5	68940	20	AAH57351	Human EST-derived
34	926.2	20.3	2616	21	AAH4335	Human Chromosome 6
35	919.6	20.1	41599	21	AAH4335	Clone 251688 of a
36	905	19.8	9339	21	AAH20962	Cosmid including s
37	905	19.8	9339	21	AAH34840	Human low adenosin
38	905	19.8	22421	21	AAH20966	Human adenosine re
39	905	19.8	22421	21	AAH34844	Human low adenosin
40	904.2	19.8	110000	22	AAH48484	Human adenosine re
41	888	19.4	5186	22	AAH84800	Nucleotide sequenc
42	839.8	18.4	138169	21	AAH43791	Human EST-derived
43	839.8	18.4	141589	21	AAH20913	Human adenosine re
44	839.8	18.4	141589	21	AAH21137	Human ELAM-1 polyn
45	839.8	18.4	141589	21	AAH21152	Human low adenosin

## ALIGNMENTS

RESULT 1	
AA129502	
ID	AA129502 standard; CDNA: 3311 BP.
AC	AA129502;
DT	12-OCT-2001 (first entry)
DE	C902P determined CDNA sequence.
XX	
KW	Human; immunotherapy; diagnosis; colon cancer; colon tumour;
KW	immunogenic; gene therapy; vaccine; colonic cancer; ss.
XX	
OS	Homo sapiens.
XX	
PN	WC200149716-A2.
PD	12-JUL-2001.
XX	
PF	29-DEC-2000; 2000WO-US35596.
XX	
PR	30-DEC-1999; 99US-0476296.
PR	10-JAN-2000; 2000US-0480321.
PR	15-FEB-2000; 2000US-0504629.
PR	06-MAR-2000; 2000US-0519444.
PR	19-MAY-2000; 2000US-0575251.
PR	29-JUN-2000; 2000US-0609448.
PR	28-AUG-2000; 2000US-0649811.
XX	
PA	(CORI-) CORIXA CORP.
XX	
PI	Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;
PI	King GE, Wang T, Jiang Y;
XX	





QY 3155 aactagagagctgtccaaaatgacagaggtttacagacatatgcttcagataaagttc 3214  
|||||  
Db 1662 aactagagagagctgtccaaaatgacagaggtttacagacatatgcttcagataaagttc 1741  
QY 3215 agacaatgagcctcatgctgtcttcttgaggcccttcacaggaatgagagctgtcttc 3274  
|||||  
Db 1742 agacaatgagcctcatgctgtcttcttgaggcccttcacaggaatgagagctgtcttc 1801  
QY 3275 agcgtccatccagcttgaagagtaaggatlaacccctccagaaagcagatgagatgaa 3334  
|||||  
Db 1802 agcgtccatccagcttgaagagtaaggatlaacccctccagaaagcagatgagatgaa 1861  
QY 3335 gcaacatgctgtgagacagccctgaggaaagagacattttcttctacccctgagaa 3394  
|||||  
Db 1862 gcaacatgctgtgagacagccctgaggaaagagacattttcttctacccctgagaa 1921  
QY 3395 tgcagcctcccaaatctctctcttgagatcccaagtgcagaaagagagagtgcttctgag 3454  
|||||  
Db 1922 tgcagcctcccaaatctctctcttgagatcccaagtgcagaaagagagagtgcttctgag 1981  
QY 3455 tggacaaaacacacaaatgagctaccccaatcccaagcattgctaggttgagctt 3514  
|||||  
Db 1982 tggacaaaacacacaaatgagctaccccaatcccaagcattgctaggttgagctt 2041  
QY 3515 ggaatatagctgtgaaagcagctacacaaacttgacccctgacttcaagctccgctgct 3574  
|||||  
Db 2042 ggaatatagctgtgaaagcagctacacaaacttgacccctgacttcaagctccgctgct 2101  
QY 3575 ccaatgtctacccctgctcccaatlaacagtgacttccaaaacgaacagagaccccaaat 3634  
|||||  
Db 2102 ccaatgtctacccctgctcccaatlaacagtgacttccaaaacgaacagagaccccaaat 2161  
QY 3635 tcccaagccctctggttagttatgcaaatatctgcgaagagcctcccaattctcaggg 3694  
|||||  
Db 2162 tcccaagccctctggttagttatgcaaatatctgcgaagagcctcccaattctcaggg 2221  
QY 3695 ccagtgacacagccctgtgtaaatcagtgaaatgaaacagcttcccttgagacacagcgg 3754  
|||||  
Db 2222 ccagtgacacagccctgtgtaaatcagtgaaatgaaacagcttcccttgagacacagcgg 2281  
QY 3755 ataattgagcagtgctgactacgaatgacagtgctgactacgaatgacttccaa 3814  
|||||  
Db 2282 ataattgagcagtgctgactacgaatgacagtgctgactacgaatgacttccaa 2341  
QY 3815 ctatgacacgaatgctgatacagtgtaaaagtgcggcctctgagagagtgtaacgag 3874  
|||||  
Db 2342 ctatgacacgaatgctgatacagtgtaaaagtgcggcctctgagagagtgtaacgag 2401  
QY 3875 ccagacgagagtgatgacccagcagagtgagcactgtacaccttgagctgagatggaga 3934  
|||||  
Db 2402 ccagacgagagtgatgacccagcagagtgagcactgtacaccttgagctgagatggaga 2461  
QY 3935 atgattgaaatacaatgaaatccacacaaagcctgaaataaataagatgatttcaacaa 3994  
|||||  
Db 2462 atgattgaaatacaatgaaatccacacaaagcctgaaataaataagatgatttcaacaa 2521  
QY 3995 agcaagtgcttcaagcagacatctcggagagtgctatttgctgcttgatgtcccaa 4054  
|||||  
Db 2522 agcaagtgcttcaagcagacatctcggagagtgctatttgctgcttgatgtcccaa 2581  
QY 4055 atgtcccatccatgcttctcccaactgagcaataccacacacttgagagggagaaatc 4114  
|||||  
Db 2582 atgtcccatccatgcttctcccaactgagcaataccacacacttgagagggagaaatc 2641  
QY 4115 acgggggagcttcataatcactgacttgagacagctccctgggagatgatatgacaaatgaa 4174  
|||||  
Db 2642 acgggggagcttcataatcactgacttgagacagctccctgggagatgatatgacaaatgaa 2701  
QY 4175 cagctccaaagtatatcttggaaataagtaacagtatcttgatctcagagagcaagttca 4234  
|||||  
Db 2702 cagctccaaagtatatcttggaaataagtaacagtatcttgatctcagagagcaagttca 2761

QY 4235 atgaatctctcaagtgaatactactgtctcctcaatcccaaggaagccaacttgagag 4294  
|||||  
Db 2762 atgaatctctcaagtgaatactactgtctcctcaatcccaaggaagccaacttgagag 2821  
QY 4295 tcttttgctttaaacaacaaacatlaacttttgaataatgagacagatcttctcatgcta 4354  
|||||  
Db 2822 tcttttgctttaaacaacaaacatlaacttttgaataatgagacagatcttctcatgcta 2881  
QY 4355 ttcaagcgtctgataagtgctgatctgaaatacgaataatcccaacttgcagagatctt 4414  
|||||  
Db 2882 ttcaagcgtctgataagtgctgatctgaaatacgaataatcccaacttgcagagatctt 2941  
QY 4415 tgttatctcccaacagagctccgacagagacacccagctcgtatgaaacgtctgctctt 4474  
|||||  
Db 2942 tgttatctcccaacagagctccgacagagacacccagctcgtatgaaacgtctgctctt 3001  
QY 4475 gtccataatcataltaacacagacacatctcgtgacatcattttaaataatgtaga 4534  
|||||  
Db 3002 gtccataatcataltaacacagacacatctcgtgacatcattttaaataatgtaga 3061  
QY 4535 agtgatagaggaactgcaactgtgtcaatagcctag 4569  
|||||  
Db 3062 agtgatagaggaactgcaactgtgtcaatagcctag 3096

RESULT 2  
AAH34879  
ID AAH34879 standard; cDNA: 2854 BP.  
XX  
AC AAH34879;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:1961.  
XX  
DE Human colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; chromosome 1; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
DR WPI; 2001-235357/24.  
DR P-PSDB; AAG75474.  
XX  
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
PS Claim 1; Page 3462-3463; 9803pp; English.  
XX  
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
CC the proteins are collectively known as colon cancer antigens. The colon  
CC cancer antigens have cytostatic activity and can be used in gene  
CC therapy and vaccine production. N and P may be used in the prevention,  
CC diagnosis and treatment of diseases associated with inappropriate P  
CC expression. For example, N and P may be used to treat disorders  
CC associated with decreased expression by rectifying mutations or deletions  
CC in a patient's genome that affect the activity of P by expressing  
CC inactive proteins or to supplement the patients own production of P.  
CC Additionally, N may be used to produce the colon cancer-associated Ps,  
CC by inserting the nucleic acids into a host cell and culturing the cell



QY	3524	gtctgcaagcaaggtcaacaaacttgaaccccttgactctcaagctcccggtgctccaaatgcta	3583
Db	1734	gtctgcgaagcaagctcccaaacacttgcaccctgacgtccacgtcccggtgctccaaatgcta	1793
QY	3584	ccctgcctcccaattacacgttgacttcccaaacgcaacaaaggaacacacgaacaaattcccagcc	3643
Db	1794	ccctgcctcccaattacacgttgacttcccaaacgcaacaaaggaacacacgaacaaattcccagcc	1853
QY	3644	ctctggtagtttatgcaaatatctgcgcaagagagctccccaattctcaaggtccagtttca	3703
Db	1854	ctctggtagtttatgcaaatatctgcgcaagagagctccccaattctcaaggtccagtttca	1913
QY	3704	cagcccgatgtgaatcgaatgaaatggaaaaacagttaaccttgaaactacgtggaataatggag	3763
Db	1914	cagcccgatgtgaatcgaatgaaatggaaaaacagttaaccttgaaactacgtggaataatggag	1973
QY	3764	caggtgtcgtatgctactaagaatgatacgtgttctactcaaggtatttcaacaattatgca	3823
Db	1974	caggtgtcgtatgctactaagaatgatacgtgttctactcaaggtatttcaacaattatgca	2033
QY	3824	cgaaatggttaagatacagtgtaaaaagttcgggctcttgggaagagttaaacgaaccaagga	3883
Db	2034	cgaaatggttaagatacagtgtaaaaagttcgggctcttgggaagagttaaacgaaccaagga	2093
QY	3884	gagtgatacccccaagcaatggtggacacgtgtacataacctgctgatttggaaatgtatgaa	3943
Db	2094	gagtgatacccccaagcaatggtggacacgtgtacataacctgctgatttggaaatgtatgaa	2153
QY	3944	tacaatggaatcccaacgaagctggaatttaataagatagttctcaacaaagaacagtgt	4003
Db	2154	tacaatggaatcccaacgaagctggaatttaataagatagttctcaacaaagaacagtgt	2213
QY	4004	gttccagagaacacatctctggggagcctcatattgtgctctgatgtgcccaaatgtctcca	4063
Db	2214	gttccagagaacacatctctggggagcctcatattgtgctctgatgtgcccaaatgtctcca	2273
QY	4064	taacctgactcttcccaacttggccaatatcaacgaaccttgaaggoggaatttccagggggca	4123
Db	2274	taacctgactcttcccaacttggccaatatcaacgaaccttgaaggoggaatttccagggggca	2333
QY	4124	gttccatataactctgacttggaaacgctctcttgggaatgattatgacatggaaacagcttca	4183
Db	2334	gttccatataactctgacttggaaacgctctcttgggaatgattatgacatggaaacagcttca	2393
QY	4184	agtatatcatctcgaaataagatacagaatttcttgatctcagagacaagtccaatgtacatcc	4243
Db	2394	agtatatcatctcgaaataagatacagaatttcttgatctcagagacaagtccaatgtacatcc	2453
QY	4244	ttcaagtgaaataactactgctctcaatcccaaggaaagccaactctgaggaagtcttcttgt	4303
Db	2454	ttcaagtgaaataactactgctctcaatcccaaggaaagccaactctgaggaagtcttcttgt	2513
QY	4304	ttaaaccgaanaaatactacttggaaaatgagcaagatcttttcaatgtatcttaacggctg	4363
Db	2514	ttaaaccgaanaaatactacttggaaaatgagcaagatcttttcaatgtatcttaacggctg	2573
QY	4364	ttgataaggtgcatactgynaatcagaaatcatcacaacttgcacagagatcatcttgttatatc	4423
Db	2574	ttgataaggtgcatactgynaatcagaaatcatcacaacttgcacagagatcatcttgttatatc	2633
QY	4424	cttcccaagaactccgcagagaacactgactgctatgaaaagctctgctcttgccttaata	4483
Db	2634	cttcccaagaactccgcagagaacactgactgctatgaaaagctctgctcttgccttaata	2693
QY	4484	ttcatatcaaacagacacattcttcgtgacattcaaatatttaaaaattatgtggaagtgtatg	4543
Db	2694	ttcatatcaaacagacacattcttcgtgacattcaaatatttaaaaattatgtggaagtgtatg	2753
QY	4544	gagaactgcaagtcttaaatgacctag 4599	
Db	2754	gagaactgcaagtcttaaatgacctag 2779	

RESULT# 3  
 AAF81787  
 ID AAF81787 standard; cDNA; 2854 BP.  
 XX  
 AC AAF81787;  
 XX  
 DT 12-JUN-2001 (first entry)  
 XX  
 DE Human secreted protein gene 1 SEQ ID NO:11.  
 XX  
 KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;  
 KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;  
 KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;  
 KW ophthalmological; neuroprotectant; nootropic; anticonvulsant; vaccine;  
 KW antialzheimer; antiparkinsonian; antimicrobial; vulnery; gene therapy;  
 KW immune disorder; hyperproliferative disorder; cardiovascular disease;  
 KW cancer; angiogenic disorder; neurological disorder; infectious disease;  
 KW wound healing; regeneration; chemotaxis; chromosome 1; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200112775-A2.  
 XX  
 PD 22-FEB-2001.  
 XX  
 PF 16-AUG-2000; 2000WO-US22325.  
 XX  
 PR 17-AUG-1999; 99US-0149182.  
 XX  
 PA (HDMA-) HDMA GENOME SCI INC;  
 XX  
 PI Rosen CA, Ni J, Florence KA, Fiscella M, Wei P, Baker RP,  
 PI Birse CE, Young PE, Komatsoulis GA, Moore PA, Soppet DR;  
 XX  
 DR WPI: 2001-147550/15.  
 P-PSDB: AAB74733.  
 XX  
 PT Nucleic acids encoding 25 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 XX  
 PS Claim 1; Page 441; 485pp; English.  
 XX  
 CC AAF81787 to AAF81817 encode the human secreted proteins given in AAB74733  
 CC to AAB74772. Human secreted proteins can have activities based on the  
 CC tissues and cells they are expressed in. Example of activities include:  
 CC immunomodulatory; antisclerotic; dermatological; immunosuppressive;  
 CC antiinflammatory; anti-HIV; immunostimulant; cytostatic; cardiant;  
 CC vascular; anti-angiogenic; ophthalmological; neuroprotectant; nootropic;  
 CC anticonvulsant; antialzheimers; antiparkinsonian; antimicrobial; and  
 CC vulnery. Human secreted proteins can be used in gene therapy and  
 CC vaccine. Human secreted protein nucleotide sequences (NAMI) and proteins  
 CC (PEPI) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate polypeptide expression. For example, NAMI  
 CC and PEPI may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patients genome  
 CC that affect the activity of proteins by expressing inactive proteins or  
 CC to supplement the patients own production of polypeptides. Disorders that  
 CC may be prevented, diagnosed and/or treated include immune disorders,  
 CC hyperproliferative disorders (e.g. cancers), cardiovascular diseases,  
 CC angiogenic disorders, neurological disorders, infectious diseases and/or  
 CC for promoting wound healing, regeneration and/or chemotaxis. AAF81778 to  
 CC AAF81786 and AAB74732 represent sequences used in the exemplification of  
 CC the present invention.  
 XX  
 SQ Sequence 2854 BP; 885 A; 628 C; 643 G; 698 T; 0 other:

QY	1544	gagaataacagggagatgtgaacgaatggggccattaaagttctgtgttaccttga	1603
Db	9	gagaaataacagggagatgtgaacgaatggggccattaaagttctgtgttaccttga	68
QY	1604	ttcttcacctcttgaagggcccttgtagtattcacttcatltagcgtgaaacaattgct	1663
Db	69	ttcttcacctcttgaagggcccttgtagtattcacttcatltagcgtgaaacaattgct	128
QY	1664	atgaagcattgtctgttgcattcgaccccaatgtgcagaagaatgaaacaacttcaac	1723
Db	129	atgaagcattgtctgttgcattcgaccccaatgtgcagaagaatgaaacaacttcaac	188
QY	1724	aaataaaggggagttacacgttcaacaagatgaagaaaggaggtcagagagaactcttc	1783
Db	189	aaata-----	193
QY	1784	ttcccccgltcaaatatatacatcacacaccacacagcacaagctcgtgtgcacaacaca	184.3
Db	194	-----	193
QY	1844	cgcccatgtcacacacagcagacatacacacacacagcacaagcactagagaagcatgtgcacc	190.3
Db	194	-----aagacatggtgacc	209
QY	1904	aggcatctctgtatctcgtcttgaagctcacaggaagcgaatttatttcaaaaatgttcca	196.3
Db	210	aggcatctctgtatctcttgaagctcacaggaagcgaatttatttcaaaaatgttcca	269
QY	1964	ttttgattccctgaaacagtgaagaacaaggctbaccatttgtagaccaaaacttggacct	202.3
Db	270	ttttgattccctgaaacagtgaagaacaaggctbaccatttgtagaccaaaacttggacct	329
QY	2024	acaaaatbctgatagttctctgtctgtagcttactctctcccaagtaatgatgaaccttca	208.3
Db	330	acaaaatbctgatagttctctgtctgtagcttactctctcccaagtaatgatgaaccttca	389
QY	2084	ctggagcagatgtggcaactgtgggagagaagggtgaaagatccactcactcctgtattca	214.3
Db	390	ctggagcagatgtggcaactgtgggagagaagggtgaaagatccactcactcctgtattca	449
QY	2144	ttgcagggaaaaaattttagctgtaatatgagcccaagaagttaggcatttgttccatgagtgg	220.3
Db	450	ttgcagggaaaaaattttagctgtaatatgagcccaagaagttaggcatttgttccatgagtgg	509
QY	2204	ctcatctacgaatggggagatatttgcacgagttacataatgatgagaatcttactatcca	226.3
Db	510	ctcatctacgaatggggagatatttgcacgagttacataatgatgagaatcttactatcca	569
QY	2264	atggaagaatatcaagcagatagaatgttccagcaggttatctctgtacaaatgtgttaaaga	232.3
Db	570	atggaagaatatcaagcagatagaatgttccagcaggttatctctgtacaaatgtgttaaaga	629
QY	2324	agttccagggagagcagttgttaccacaaaagaatgacacattcaataagttaacagagacct	238.3
Db	630	agttccagggagagcagttgttaccacaaaagaatgacacattcaataagttaacagagacct	689
QY	2384	atgaaaaaaggatgtgagtttctgttcccaalcccccgcagacggaagaaggtctctataatgt	244.3
Db	690	atgaaaaaaggatgtgagtttctgttcccaalcccccgcagacggaagaaggtctctataatgt	749
QY	2444	ttgcacaacaatgttgtattctataatgttgtaattctctgtacagaacaaaacacaagaagaag	250.3
Db	750	ttgcacaacaatgttgtattctataatgttgtaattctctgtacagaacaaaacacaagaagaag	809
QY	2504	ctccaaaacaagcaaaaatcaaaaatgtcaaatctcccgaaagcagatggaagtgttccgtgatt	256.3
Db	810	ctccaaaacaagcaaaaatcaaaaatgtcaaatctcccgaaagcagatggaagtgttccgtgatt	869
QY	2564	ctggagacctttaaagaaaacacactccttatgacacacacagccacaaatcccaacttccat	262.3
Db	870	ctggagacctttaaagaaaacacactccttatgacacacacagccacaaatcccaacttccat	929
QY	2624	tgtctgcagattggaacaagaattgtgtttagtcccttgtaacaaatctggaagcatgtgcga	268.3

Db	930	tcgtcagatctgagcaaaagattgtgtgtttagtctcttgacaatctggaaacatgtagcga	989
Qy	2684	cttgtaacgcctccaatctgcagtaactcaacgaagccagctcttccctgcgcagacattg	2743
Db	990	ctgtgaacgcctccaatctgcagtaactcaacgaagccagctcttccctgcgcagacattg	1049
Qy	2744	agctggggtccttggtgttgagctgtgacatttgacagctgtctccatgctacaaaatgac	2803
Db	1050	agctggggtccttggtgttgagctgtgacatttgacagctgtctccatgctacaaaatgac	1109
Qy	2804	tcataagaataaacagttgacgttgacgttgacagggacacactgcgcaaaagattacctgaacg	2863
Db	1110	tcataagaataaacagttgacgttgacagggacacacactgcgcaaaagattacctgaacg	1169
Qy	2864	cttcaggaaggagctcccaactctgcagcgggcttcgattccgcatcttcatatgtgcaac	2923
Db	1170	cttcaggaaggagctcccaactctgcagcgggcttcgattccgcatcttcatatgtgcaac	1216
Qy	2924	atttgcctgttttcatagacacacagcagttatgtggagttgcgaagaagaaatccaaatl	2983
Db	1217	-----	1216
Qy	2984	gggctctctgctgcagcttagtgatatgaagaataatccaaactgatatgaa	3043
Db	1217	-----ggtatgaagaataatccaaactgatatgaa	1253
Qy	3044	ttgtcctgtctgcaggtatggggaggaacaaactatagttgggtctttagcagggccaac	3103
Db	1254	ttgtcctgtctgcaggtatggggaggaacaaactatagttgggtctttagcagggccaac	1313
Qy	3104	aaagtgtgtccaatccacacagctgcgtcttggggccctctgcagctcaagaactagag	3163
Db	1314	aaagtgtgtccaatccacacagctgcgtcttggggccctctgcagctcaagaactagag	1373
Qy	3164	agctgtccaaaatgacaggaagtttacagacatatgctcagatcaagttcaagaacaatg	3223
Db	1374	agctgtccaaaatgacaggaagtttacagacatatgctcagatcaagttcaagaacaatg	1433
Qy	3224	gctcaatgatcttcttggggcccttcaacgaagaaatgaaactgtctcagcgtcca	3283
Db	1434	gctcaatgatcttcttggggcccttcaacgaagaaatgaaactgtctcagcgtcca	1493
Qy	3284	tccagcttgagagtaagggatcaacctccctccagaaacagccagctggaatgtagcagatga	3343
Db	1494	tccagcttgagagtaagggatcaacctccctccagaaacagccagctggaatgtagcagatga	1553
Qy	3344	tcgttgacagcacgttgggaaagacacttgtttcttatcaactgcagacaatgacgctc	3403
Db	1554	tcgttgacagcacgttgggaaagacacttgtttcttatcaactgcagacaatgacgctc	1613
Qy	3404	cccaatccctctctcggatcccaatggaacaaagcaaggtggtctttagtggacaata	3463
Db	1614	cccaatccctctctcggatcccaatggaacaaagcaaggtggtctttagtggacaata	1673
Qy	3464	acaccaaaatggctactcccaatcccaagcattgtctaaagtttgacacttggaaataca	3523
Db	1674	acaccaaaatggctactcccaatcccaagcattgtctaaagtttgacacttggaaataca	1733
Qy	3524	gtctgcaagcaagctcaacaacttgcacctgactgtcaagtcccgtycgtccaatgtcta	3583
Db	1734	gtctgcaagcaagctcaacaacttgcacctgactgtcaagtcccgtycgtccaatgtcta	1793
Qy	3584	ccctgcctcccaattacagttgacttccaaaacgaacaaagaaacccgcaaatctcccaacc	3643
Db	1794	ccctgcctcccaattacagttgacttccaaaacgaacaaagaaacccgcaaatctcccaacc	1853
Qy	3644	ctctgtagattatgacaatatctgcgaagagacgtccccaattctcagggccaatgtgca	3703
Db	1854	ctctgtagattatgacaatatctgcgaagagacgtccccaattctcagggccaatgtgca	1913
Qy	3704	cagccctgattgaatcagttgaatggaaaaacagttaccttggaaactatgataatggag	3763

Db	1914	cagcccttgatcgatcagtgtaatggaataaacagttaaccttggaactctgtaataatggag	1973
Qy	3764	caggtgcgcgatacgtactaaagataagcgggtgtctactcaaggatattacaactatgaca	3823
Db	1974	cagttgcgcgatacgtactaaagataagcgggtgtctactcaaggatattacaactatgaca	2033
Qy	3824	cgaatgtgtagatacagtgtaaaagtctcggtgctctcgtggagagatgtaacgcagccagacgga	3883
Db	2034	cgaatgtgtagatacagtgtaaaagtctcggtgctctcgtggagagatgtaacgcagccagacgga	2093
Qy	3884	gagtgatcccccaacagcgttggaacactgtatatacacttgctgtgattggaataatgataaa	3943
Db	2094	gagtgatcccccaacagcgttggaacactgtatatacacttgctgtgattggaataatgataaa	2153
Qy	3944	tacatgtaatcccaacaaagacctataaataagataatgataatgtttcaacacacaaagatgt	4003
Db	2154	tacatgtaatcccaacaaagacctataaataagataatgataatgtttcaacacacaaagatgt	2213
Qy	4004	gtttcagcagaacacatcctcgggaggctcaatttgttgctctgatagtgtcccaatgtccca	4063
Db	2214	gtttcagcagaacacatcctcgggaggctcaatttgttgctctgatagtgtcccaatgtccca	2273
Qy	4064	tactctgactcttcccaactcgtgccaaataacccagccctggaagcgggaattccaacggggca	4123
Db	2274	tactctgactcttcccaactcgtgccaaataacccagccctggaagcgggaattccaacggggca	2333
Qy	4124	gtctcatatactgacttggaacagctcctcgtgggatagtttatgacacatgggaacagctaca	4183
Db	2334	gtctcatatactgacttggaacagctcctcgtgggatagtttatgacacatgggaacagctaca	2393
Qy	4184	agtatcatctcgaataagtaagtaacagatctcttgatactcagagacaagttcaatgatatctc	4243
Db	2394	agtatcatctcgaataagtaagtaacagatctcttgatactcagagacaagttcaatgatatctc	2453
Qy	4244	ttcaagtgataatactactgctctcaatcccaagaagaccacactctggaagaagtccttttgt	4303
Db	2454	ttcaagtgataatactactgctctcaatcccaagaagaccacactctggaagaagtccttttgt	2513
Qy	4304	ttaaaccggaataaacattccttttgaataatggcacagatcctttatctgcatcattcagctg	4363
Db	2514	ttaaaccggaataaacattccttttgaataatggcacagatcctttatctgcatcattcagctg	2573
Qy	4364	ttgataaggtcgatctgtaaacatcaagaatatccaacatctgcagataccttggttattc	4423
Db	2574	ttgataaggtcgatctgtaaacatcaagaatatccaacatctgcagataccttggttattc	2633
Qy	4424	ctccaacagatccgcgcagagacacctagctctgtatgaacgctctgctctgttccataa	4483
Db	2634	ctccaacagatccgcgcagagacacctagctctgtatgaacgctctgctctgttccataa	2693
Qy	4484	ttcatatcaagaagacacctctctctgcatcattcaatlttaaaaaataatgttgaaatggatag	4543
Db	2694	ttcatatcaagaagacacctctctctgcatcattcaatlttaaaaaataatgttgaaatggatag	2753
Qy	4544	gagaactcgacgtgcaataagcttag	4569
Db	2754	gagaactcgacgtgcaataagcttag	2779
RESULT 4			
AAZ09840			
ID	AAZ09840 standard; cDNA, 3111 BP.		
XX	AAZ09840;		
AC			
XX			
DT	26-NOV-1999 (first entry)		
XX			
DE	Human membrane spanning protein MSP-5 cDNA fragment 2.		
XX			
KW	Membrane spanning protein; MSP; human; treatment; diagnosis; prevention; neoplastic disorder; immunological disorder; reproductive disorder; MSP-5; ds.		
XX			

OS Homo sapiens.  
XX  
PN MO9946380-A2.  
XX  
PD 16-SEP-1999.  
XX  
PF 09-MAR-1999; 99MO-USO5073.  
XX  
PR 13-MAR-1998; 98US-0039064.  
XX  
PA (INCYTE) INCYTE PHARM INC.  
XX  
PI Tang YT, Bandman O, Lal P, Hillman JL, Yue H, Corley NC;  
PI Guegler KJ, Kaser MR, Baughn MR, Shah P;  
XX  
DR MPI; 1999-551409/46.  
DR P-PSDB; AAV33298.  
XX  
PT New human membrane spanning proteins used to, e.g. prevent and treat  
PT neoplastic disorders -  
XX  
PS Example 1: Page 80-81; 81pp; English.  
XX  
CC This invention describes novel human membrane spanning proteins (MSPs),  
CC and the polynucleotides encoding them. The products of the invention are  
CC used to diagnose, prevent and treat neoplastic, immunological and  
CC reproductive disorders. This sequence encodes a human membrane spanning  
CC protein MSP-5 fragment.  
XX  
SQ Sequence 3111 BP; 968 A; 664 C; 691 G; 788 T; 0 other;

[illegible]

```
|||||
Db 391 gagcagatgaggcaactctgagagaaaggtgtaaaggtccaccctcaccctcggatccaatt 450
2146 gcagaaaaaagttagctgaaatacgagccacaagaagtaggcatctgtccatgagtgagct 2205
451 gcagaaaaaagttagctgaaatacgagccacaagaagtaggcatctgtccatgagtgagct 510
2206 catctacgattgaggagatttgacgagtaacaatacgatgagaattctacttaccat 2265
511 catctacgattgaggagatttgacgagtaacaatacgatgagaattctacttaccat 570
2266 ggaagaatacagaagcagtaagaatggttcagcgagttactcgtgtacaataatgtgttaagaag 2325
571 ggaagaatacagaagcagtaagaatggttcagcgagttactcgtgtacaataatgtgttaagaag 630
2326 tgtcaggagagcgacgtcttaacccaagaagtcacatccaataaagtaacagagcctcat 2385
631 tgtcaggagagcgacgtcttaacccaagaagtcacatccaataaagtaacagagcctcat 690
2386 gaaaaagatgtagagttgttctccaaatcccgccagacgagaaagcctctataatgtt 2445
691 gaaaaagatgtagagttgttctccaaatcccgccagacgagaaagcctctataatgtt 750
2446 gcaacaacatgtgttctctataatgttgaattctgttacaacaacaacccaacaagaagagct 2505
751 gcaacaacatgtgttctctataatgttgaattctgttacaacaacaacccaacaagaagagct 810
2506 ccaacaagcaaaaatcacaataatgcaatctccgaagcacaatggagaagtcagatctct 2565
811 ccaacaagcaaaaatcacaataatgcaatctccgaagcacaatggagaagtcagatctct 870
2566 gaggactttaagaaaacactcctctatgacaacacagccacaacatcccaacttctcatgt 2625
871 gaggactttaagaaaacactcctctatgacaacacagccacaacatcccaacttctcatgt 930
2626 ctgcaagtttgacaacaagaatgtgtgtttagtccttgcacaacaatctggaagatgagcact 2685
931 ctgcaagtttgacaacaagaatgtgtgtttagtccttgcacaacaatctggaagatgagcact 990
2686 ggtacccgctccaatcgactgaaatcagaagcagcgagctttccctcgctgcagacagttgag 2745
991 ggtacccgctccaatcgactgaaatcagaagcagcgagctttccctcgctgcagacagttgag 1050
2746 ctggggtctctgggttgagatgagtgacattgacagtgctcgcccatgtacaataatgaaatc 2805
1051 ctggggtctctgggttgagatgagtgacattgacagtgctcgcccatgtacaataatgaaatc 1110
2806 atcacagataaacagtgagtgagtgagagagacacactgcgcaaaagattacctgcagcagct 2865
1111 atcacagataaacagtgagtgagtgagagagacacactgcgcaaaagattacctgcagcagct 1170
2866 tcaaggagagcgtccatctgcagcgagcttcgatccgcatcttaactatgtgacaacat 2925
1171 tcaaggagagcgtccatctgcagcgagcttcgatccgcatcttaactatgtgacaacat 1215
2926 ttgcctgttttccatgacacacagcagttatgaggagtgcgacaagaanaalcccaattgg 2985
1216 ----- 1215
2986 ggcctcttgccctgcagcctagtgattagagaagaatatcccaactgtagtctgaaatc 3045
1216 -----gtagttagagaagaatatcccaactgtagtctgaaatc 1254
3046 gtgcctgtgacggaatggggaagacaacacataaagtgggtcttaacgaggtacaacaa 3105
1255 gtgcctgtgacggaatggggaagacaacacataaagtgggtcttaacgaggtacaacaa 1314
3106 agtgggtgcatcatccacacagtcgcttgaggccctctgcagctcaagaactagaagag 3165
1315 agtgggtgcatcatccacacagtcgcttgaggccctctgcagctcaagaactagaagag 1374
3166 ctgtcccaaatgacaggaaggtttacagacatagtctccagatcaagtctcagaacaatgagc 3225
|||||
```

---

```
Db 1375 ctgtccaaaatgacaggaaggtttacagacatagtcttcacagatccaaagtctcagaacaatgagc 1434
3226 ctcatgtatgcttttgaggcccttctcaaggaatgagagctgtctccagcgtccatc 3285
1435 ctcatgtatgcttttgaggcccttctcaaggaatgagagctgtctccagcgtccatc 1494
3286 cagcttgagagtaaggagttaaacccctccagacaagcagtgagatgaaatgagacagtgatc 3345
1495 cagcttgagagtaaggagttaaacccctccagacaagcagtgagatgaaatgagacagtgatc 1554
3346 gtgagacagcccggtgggaagaggaactgtgttcttcaactcctgagacaatgagccctccc 3405
1555 gtgagacagcccggtgggaagaggaactgtgttcttcaactcctgagacaagcagcctccc 1614
3406 caaatcctctctgtggatcccaagtgagacaagaagaagtgagcttctgagtgagacaanaac 3465
1615 caaatcctctctgtggatcccaagtgagacaagaagaagtgagcttctgagtgagacaanaac 1674
3466 accaaaatggcctacccctccaaatcccaagcattgtctaagttggcacttggaaatacagt 3525
1675 accaaaatggcctacccctccaaatcccaagcattgtctaagttggcacttggaaatacagt 1734
3526 ctggaagaagcttcaacaacttgcacctgactgttcaagttcccggtggtccatgtctacc 3585
1735 ctggaagaagcttcaacaacttgcacctgactgttcaagttcccggtggtccatgtctacc 1794
3586 ctgctcccaattacagtgacttccaaaagcaagaagacaaccaaatccccaagcct 3645
1795 ctgctcccaattacagtgacttccaaaagcaagaagacaaccaaatccccaagcct 1854
3646 ctggttagttatgcaaatatccgcaagagagctccccaattctcaaggccagtggttaca 3705
1855 ctggttagttatgcaaatatccgcaagagagctccccaattctcaaggccagtggttaca 1914
3706 gccctgattgaatcgaatggaatggaataacagttacacttgaactcggatgaggaagca 3765
1915 gccctgattgaatcgaatggaatggaataacagttacacttgaactcggatgaggaagca 1974
3766 ggtgctatgcttactaaagtgacggtgtctactcaagagatttcaacaacttataagacag 3825
1975 ggtgctatgcttactaaagtgacggtgtctactcaagagatttcaacaacttataagacag 2034
3826 aatgtatgatacagtgtaaaagtgcgggctctgggagaggttaaagcagccagaagagaga 3885
2035 aatgtatgatacagtgtaaaagtgcgggctctgggagaggttaaagcagccagaagagaga 2094
3886 gtgtgaccccaagcagatgtagacactgtatataactgtgcggatttgagaaatgtagaata 3945
2095 gtgtgaccccaagcagatgtagacactgtatataactgtgcggatttgagaaatgtagaata 2154
3946 caatggaatccacaagaagcctgaaatcaataagagatgatttcaacaagaagcagtggt 4005
2155 caatggaatccacaagaagcctgaaatcaataagagatgatttcaacaagaagcagtggt 2214
4006 ttcaagcaaatccctcggagagcctcaattgtggtctctgagtggtcccaatgtgccata 4065
2215 ttcaagcaaatccctcggagagcctcaattgtggtctctgagtggtcccaatgtgccata 2274
4066 cctgattcttcccaacttgcgccaatccacagcactgaaagcgggaatttcaacggggagcagc 4125
2275 cctgattcttcccaacttgcgccaatccacagcactgaaagcgggaatttcaacggggagcagc 2334
4126 ctcaataatctgacttgagacagctccctgggagatgattatataccttggaacagcctcaacg 4185
2335 ctcaataatctgacttgagacagctccctgggagatgattatataccttggaacagcctcaacg 2394
4186 tataatctcgaataaagtaacaagtatcttgcattccagagacaagtccaatgaaatcctt 4245
2395 tataatctcgaataaagtaacaagtatcttgcattccagagacaagtccaatgaaatcctt 2454
4246 caagtgtataactactgtcttccatcccaagaagaagcacaactctgaagagagcttttgtt 4305
2455 caagtgtataactactgtcttccatcccaagaagaagcacaactctgaagagagcttttgtt 2514
Db
```





Dh 742 acaaatgttattctatagtattgaattcttctgtacagaacaaacacaaagaagctcc 801  
QY 2508 aaacaagcaaaaatcaaaaatgcgaatctcccgagacatctgggaagtgatccgttgattctga 2567  
Dh 802 aaacaagcaaaaatcaaaaatgcgaatctcccgagacatctgggaagtgatccgttgattctcga 861  
QY 2568 ggaatttaagaanaaacccactctatgtacaacacagccacccaatccacattctcatgtct 2627  
Dh 862 ggaatttaagaanaaacccactctatgtacaacacagccacccaatccacattctcatgtct 921  
QY 2628 gcaagattggacaagaagaatttggttgatttgattcttgacaacacttggaagcatggagcagcg 2687  
Dh 922 gcaagattggacaagaagaatttggttgatttgattcttgacaacacttggaagcatggagcagcg 981  
QY 2688 taacgcctcactcactgcactgtgaatccaagcagcgccattctctctgctgcagacagttgagct 2747  
Dh 982 taacgcctcactcactgcactgtgaatccaagcagcgccattctctctgctgcagacagttgagct 1041  
QY 2748 ggggtctctgggttggtgattggtgacatttgacagtgcctgcgcacatgtacaaaatgaactcat 2807  
Dh 1042 ggggtctctgggttggtgattggtgacatttgacagtgcctgcgcacatgtacaaaatgaactcat 1101  
QY 2808 acagataaacagcttgacagtgacaggaacacactgcgcaaaaagattacacttgacagcagcttc 2867  
Dh 1102 acagataaacagcttgacagtgacaggaacacactgcgcaaaaagattacacttgacagcagcttc 1161  
QY 2868 agagaggagacgtccactctctgcagcggcttcgactcgacatttactgtataltgttgacaacttt 2927  
Dh 1162 agagaggagacgtccactctctgcagcggcttcgactcgacatttactgtataltgttgacaacttt 1204  
QY 2928 ggcctgtttccatgtacaacacagcagttatggggaagtgcgcagagaanaatcccaattgggc 2987  
Dh 1205 ----- 1204  
QY 2988 ctctcgtcgccgcagccttagtgaattgaagaanaatatcccaactgatgtatgaataattgt 3047  
Dh 1205 -----ggaattgaagaanaatatcccaactgaatggaatcttgaaattgt 1245  
QY 3048 gctctgacagtgatggggaagaacacacactataagtggttgctttaacagaggtcaacaacaag 3107  
Dh 1246 gctctgacagtgatggggaagaacacacactataagtggttgctttaacagaggtcaacaacaag 1305  
QY 3108 tgggtgcacatcatccacacagtcgctttggggccctctgcagctcaagaataagagagagct 3167  
Dh 1306 tgggtgcacatcatccacacagtcgctttggggccctctgcagctcaagaataagagagagct 1365  
QY 3168 gtccaaatatgacagagaggtttacagacatatgtctcagatcaagaattagaacaatgtgcct 3227  
Dh 1366 gtccaaatatgacagagaggtttacagacatatgtctcagatcaagaattagaacaatgtgcct 1425  
QY 3228 catgtatgcttttggggccctttcatcaggaatlgagctgtctctcagcgtccatcca 3287  
Dh 1426 catgtatgcttttggggccctttcatcaggaatlgagctgtctctcagcgtccatcca 1485  
QY 3288 gcttgagagtgaaaggtttaacacctcagaacagccagcttgatgaatggcacagtgtatcgt 3347  
Dh 1486 gcttgagagtgaaaggtttaacacctcagaacagccagcttgatgaatggcacagtgtatcgt 1545  
QY 3348 ggaagacacggttggaagaagacacttgctttctatcatcccggaacatgacgtccctccca 3407  
Dh 1546 ggaagacacggttggaagaagacacttgctttctatcatcccggaacatgacgtccctccca 1605  
QY 3408 aatcctctctcgtgatacccgctgagacagaagcaagtgctgttgattagtggaacaaacac 3467  
Dh 1606 aatcctctctcgtgatacccgctgagacagaagcaagtgctgttgattagtggaacaaacac 1665  
QY 3468 caaatggtcctactccaatatcccaaggcatgttaagtttgacacttggaatatcagctct 3527  
Dh 1666 caaatggtcctactccaatatcccaaggcatgttaagtttgacacttggaatatcagctct 1725  
QY 3528 gcaagcagagctcaaaaacacttgacccctgactgtcaagctcccgctgctcccaatgtaccct 3587  
Dh 1726 gcaagcagagctcaaaaacacttgacccctgactgtcaagctcccgctgctcccaatgtaccct 1785

QY 3588 gctccaattacagtgacttccaaaacgaaacagagacacagcaaatctcccaagccctct 3647  
Dh 1786 gctccaattacagtgacttccaaaacgaaacagagacacagcaaatctcccaagccctct 1845  
QY 3648 ggtatgtttatgcaaatattctgcgaagagagctccccaattctcagggcagtgatcacagc 3707  
Dh 1846 ggtatgtttatgcaaatattctgcgaagagagctccccaattctcagggcagtgatcacagc 1905  
QY 3708 ccttgattgaatcaatgtgaatgaaanaacagttactcttggaactacatgtatgtagagcag 3767  
Dh 1906 ccttgattgaatcaatgtgaatgaaanaacagttactcttggaactacatgtatgtagagcag 1965  
QY 3768 tgcctgacttaactaaagatgacggtgtctactactcaaggattattccaactatgacacgaa 3827  
Dh 1966 tgcctgacttaactaaagatgacggtgtctactactcaaggattattccaactatgacacgaa 2025  
QY 3828 tggtagatatacagtgataaagtgcgggtctctgggaggggttaacgcgaacccagaggaagt 3887  
Dh 2026 tggtagatatacagtgataaagtgcgggtctctgggaggggttaacgcgaacccagaggaagt 2085  
QY 3888 gatccccagcagagtgagcactgtacatacctgtgcttggaatgagatgtatgaatatca 3947  
Dh 2086 gatccccagcagagtgagcactgtacatacctgtgcttggaatgagatgtatgaatatca 2145  
QY 3948 atggaatcccaaccaagaccttgaanaattaaagatgatatgttcaacacagaagtggttt 4007  
Dh 2146 atggaatcccaaccaagaccttgaanaattaaagatgatatgttcaacacagaagtggttt 2205  
QY 4008 cagcagaacaatccctcggagagctcatcttggtctctcgatggtcccaatgtcccatacc 4067  
Dh 2206 cagcagaacaatccctcggagagctcatcttggtctctcgatggtcccaatgtcccatacc 2265  
QY 4068 tgaatcttccacacttggtcccaatcacccgagcttgaagcggaaattcaacggggagctct 4127  
Dh 2266 tgaatcttccacacttggtcccaatcacccgagcttgaagcggaaattcaacggggagctct 2225  
QY 4128 catatactgacttggacagctccctggggaatgataatgacacatlgaaacagctcaacaagta 4187  
Dh 2226 catatactgacttggacagctccctggggaatgataatgacacatlgaaacagctcaacaagta 2285  
QY 4188 tatcatctgaaataagtgacaagtatctcttgatctccagagacaagtgtcaatctcttca 4247  
Dh 2286 tatcatctgaaataagtgacaagtatctcttgatctccagagacaagtgtcaatctcttca 2445  
QY 4248 agtgaatacactgtctctcatcccaaggagccaaactctgaggaatctctttgtttaa 4307  
Dh 2446 agtgaatacactgtctctcatcccaaggagccaaactctgaggaatctctttgtttaa 2505  
QY 4308 accagaanaacattacttttgaanaatggcacagatcttttcaatgtctatccaagcgtttga 4367  
Dh 2506 accagaanaacattacttttgaanaatggcacagatcttttcaatgtctatccaagcgtttga 2565  
QY 4368 taaggtcgtatctgaaatcaagaataataccaacattgacgcaggtatctttgttatctctcc 4427  
Dh 2566 taaggtcgtatctgaaatcaagaataataccaacattgacgcaggtatctttgttatctctcc 2625  
QY 4428 acagactccgcgaagacactatgctctgataagagctgcgctcctgtccaaatatcca 4487  
Dh 2626 acagactccgcgaagacactatgctctgataagagctgcgctcctgtccaaatatcca 2665  
QY 4488 tatcaacagacacatctcttgcatctcacatttaaaaaatlatgtggaagtgtgataagaga 4547  
Dh 2686 tatcaacagacacatctcttgcatctcacatttaaaaaatlatgtggaagtgtgataagaga 2745  
QY 4548 actgacagctgtcaatagcctag 4569  
Dh 2746 actgacagctgtcaatagcctag 2767

RESULT 6  
AAH33285 standard; cDNA; 2867 BP.  
ID AAH33285



XX AAH33285;  
 AC  
 XX 03-SEP-2001 (first entry)  
 DT  
 XX  
 DE Human colon cancer antigen encoding cDNA SEQ ID NO:341.  
 XX  
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma; chromosome 1; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200122920-A2.  
 PN  
 XX 05-APR-2001.  
 PD  
 XX 28-SEP-2000; 2000WO-US26524.  
 PF  
 XX 29-SEP-1999; 99US-0157137.  
 PR  
 XX 03-NOV-1999; 99US-0163280.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ruben SM, Barash SC, Birse CE, Rosen CA;  
 XX  
 DR WPI; 2001-235357/24.  
 DR P-PSDB: AAG73854.  
 XX  
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 XX  
 PS Claim 1; Page 2452-2453; 9803pp; English.  
 XX  
 CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing P.  
 CC Inactive proteins or to supplement the patients own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated Ps,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAG77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX  
 SQ Sequence 2867 BP; 891 A; 633 C; 643 G; 700 T; 0 other;

Query Match 54.2%; Score 2477; DB 22; Length 2867;  
 Best Local Similarity 91.4%; Pred No. 0;  
 Matches 2766; Conservative 0; Mismatches 5; Indels 256; Gaps 3;

OY 1544 gagaatcacagagagatgtacagcaatggggccatttaagaattctgttcatcttga 1603  
 DB 12 gagaatcacagagagatgtacagcaatggggccatttaagaattctgttcatcttga 71  
 OY 1604 ttcttcaactttagaaggagccctgagttaattcaattcagctggaacaatagct 1663  
 DB 72 ttcttcaactttagaaggagccctgagttaattcaattcagctggaacaatagct 131  
 OY 1664 atgaagagcatgtcgttgaatcgaccccaatgtgcagaagaatgaacaactcatcaac 1723  
 DB 132 atgaagagcatgtcgttgaatcgaccccaatgtgcagaagaatgaacaactcatcaac 191  
 OY 1724 aataaaggggagtagacgctcacaagaatgtagaaggagagtcagagagaactctctc 1783

DB 192 aata----- 196  
 OY 1784 ttcccccgttcaatatatacatatacacacacacacacagcagcagctggtgtgcacacaca 1843  
 DB 197 ----- 196  
 OY 1844 cgcacatgcacacagcagacatatacgcgcacacacacagcagcagctagaagagcagtgacc 1903  
 DB 197 -----aagacatggtgacc 212  
 OY 1904 aggcattctgtatctgtcttgaagctacaggaagcgatattatccaataatgttgcca 1963  
 DB 213 aggcattctgtatctgtcttgaagctacaggaagcgatattatccaataatgttgcca 272  
 OY 1964 ttcttattccctgaacaatgaagaacagagctgactatgttgagaccacaacttgagacct 2023  
 DB 273 ttcttattccctgaacaatgaagaacagagctgactatgttgagaccacaacttgagacct 332  
 OY 2024 acaaaaatgctgatttctgtgtctgagcttactctccaggttaattgaaccctaca 2083  
 DB 333 acaaaaatgctgatttctgtgtctgagcttactctccaggttaattgaaccctaca 392  
 OY 2084 ctgagcagaatgggcaactgttgagagaagggtgaagaatccacctactctgatttca 2143  
 DB 393 ctgagcagaatgggcaactgttgagagaagggtgaagaatccacctactctgatttca 452  
 OY 2144 ttgcaggaaaaaagttaagctgaatatgacccacaaggtgagcattgttccatgagtgg 2203  
 DB 453 ttgcaggaaaaaagttaagctgaatatgacccacaaggtgagcattgttccatgagtgg 512  
 OY 2204 ctcatctacgattgggagagattttgacgagttacatatatgatagaattctacttcca 2263  
 DB 513 ctcatctacgattgggagagattttgacgagttacatatatgatagaattctacttcca 572  
 OY 2264 atggaagaatagaagcagtaagaatgttccagcaggtatctgtgtacaatgtagtaaga 2323  
 DB 573 atggaagaatagaagcagtaagaatgttccagcaggtatctgtgtacaatgtagtaaga 632  
 OY 2324 agtgtcaaggagagcagctgtttacccaagaatgacatccaataagtaacagagctct 2383  
 DB 633 agtgtcaaggagagcagctgtttacccaagaatgacatccaataagtaacagagctct 692  
 OY 2384 atgaaagaatgtgagctgttctccaatcccgacagcgggaagcttcataatgt 2443  
 DB 693 atgaaagaatgtgagctgttctccaatcccgacagcgggaagcttcataatgt 752  
 OY 2444 ttgcacaacatgttgaattctataatgtgaattctgttacaagaacaaacacaagaagaag 2503  
 DB 753 ttgcacaacatgttgaattctataatgtgaattctgttacaagaacaaacacaagaagaag 812  
 OY 2504 ctccaacaagaacaaatccaataatgtcaatctccgaagcagatggggaagtgatcgagatt 2563  
 DB 813 ctccaacaagaacaaatccaataatgtcaatctccgaagcagatggggaagtgatcgagatt 872  
 OY 2564 ctgagagctttaagaagaacacactccatgtacaacacagcagccaatcccaacttccat 2623  
 DB 873 ctgagagctttaagaagaacacactccatgtacaacacagcagccaatcccaacttccat 932  
 OY 2624 tgcctgcagattgacaagaagaattgtgtttagtctcttgaacaatctggaagcatggcga 2683  
 DB 933 tgcctgcagattgacaagaagaattgtgtttagtctcttgaacaatctggaagcatggcga 992  
 OY 2684 ctggttaacccgctcaatgcagtaagaatccaagcagcgagcttttccgtgcgagcagtg 2743  
 DB 993 ctggttaacccgctcaatgcagtaagaatccaagcagcgagcttttccgtgcgagcagtg 1052  
 OY 2744 agctgggtccctgggttgggatgtgtacatttgaacgtgctgcacatgtacaagaatgaac 2803  
 DB 1053 agctgggtccctgggttgggatgtgtacatttgaacgtgctgcacatgtacaagaatgaac 1112  
 OY 2804 tcatagataaacaagt-ggcagtgacagggacacactgcgaagaagaattactctgacga 2862

Db 1113 tcatacagataaacaagctgagctgacagggagacacactgcgcaaaagattactctgcagca 1172  
 QY 2863 gcttcaagaggaagctccatctgcagcgggttcgatccgattactatgtaagca 2922  
 Db 1173 gcttcaagaggaagctccatctgcagcgggttcgatccgattactatgtaagca 2922  
 QY 2923 catttgcctgtttccatgacacacagcattatggtgagctgcgaacaagaataatccaaat 2982  
 Db 1221 ----- 1220  
 QY 2983 tgggctctctcgcgtgcagcttagtttaggaagaataatccaaatgattgtaaa 3042  
 Db 1221 -----gtgattaggaagaataatccaaatgattgtaaa 1256  
 QY 3043 attgtcgtcgtacagctgagtgaggaaagacacacataaagtgtgctttaaagagtcaca 3102  
 Db 1257 attgtcgtcgtacagctgagtgaggaaagacacacataaagtgtgctttaaagagtcaca 1316  
 QY 3103 caaagtgtgacatccatccacacacagctcgttggggcccttcgagctcaagaataagag 3162  
 Db 1317 caaagtgtgacatccatccacacacagctcgttggggcccttcgagctcaagaataagag 1376  
 QY 3163 gagctgtcccaaaatgacagaggggtttacagacatagcttcaagatcaagtccagaacat 3222  
 Db 1377 gagctgtcccaaaatgacagaggggtttacagacatagcttcaagatcaagtccagaacat 1436  
 QY 3223 ggcctcattgactgtcttctggggcccttcacaggaatgagctgtccacagctcc 3282  
 Db 1437 ggcctcattgactgtcttctggggcccttcacaggaatgagctgtccacagctcc 1496  
 QY 3283 atccagcttgagagtaagggatlaaacctccagaacagccagtgatgaatgagcaagtg 3342  
 Db 1497 atccagcttgagagtaagggatlaaacctccagaacagccagtgatgaatgagcaagtg 1556  
 QY 3343 atcgttgacaacacacgttgggaaggaacacatttcttcaacattgacagctcagcct 3402  
 Db 1557 atcgttgacaacacacgttgggaaggaacacatttcttcaacattgacagctcagcct 1616  
 QY 3403 ccccaaatcctctctgtggatccacgtgagacagagcaagtggtctgtgtagtgacaaa 3462  
 Db 1617 ccccaaatcctctctgtggatccacgtgagacagagcaagtggtctgtgtagtgacaaa 1676  
 QY 3463 aacacaaaatgacctactcccaatcccaagcattgctaaagttgacacttggaataac 3522  
 Db 1677 aacacaaaatgacctactcccaatcccaagcattgctaaagttgacacttggaataac 1736  
 QY 3523 agtctgacaagaagctacaacaccttgaacctgaactgacagctccggtgcctcaatgct 3582  
 Db 1737 agtctgacaagaagctacaacaccttgaacctgacagctccggtgcctcaatgct 1796  
 QY 3583 accctgctccaattacagtgacttccaaaacgacaaagacacagcaaatcccccagc 3642  
 Db 1797 accctgctccaattacagtgacttccaaaacgacaaagacacagcaaatcccccagc 1856  
 QY 3643 cctctgtgagtattatgcaataatccgcaagagcctcccaattccagggccagtgct 3702  
 Db 1857 cctctgtgagtattatgcaataatccgcaagagcctcccaattccagggccagtgct 1916  
 QY 3703 acaacccctgattgacagtgatgaatggaanaacagttaccttggagactactgtaagtgga 3762  
 Db 1917 acaacccctgattgacagtgatgaatggaanaacagttaccttggagactactgtaagtgga 1976  
 QY 3763 gcaagtgtcgtactgactaagatgacggtgtctactcaagaatlttcaacaactatgac 3822  
 Db 1977 gcaagtgtcgtactgactaagatgacggtgtctactcaagaatlttcaacaactatgac 2036  
 QY 3823 acgaatgtgatacagtgatgaatggaanaagtcgggctctgggagagtgatgaacgagcagaag 3882  
 Db 2037 acgaatgtgatacagtgatgaatggaanaagtcgggctctgggagagtgatgaacgagcagaag 2096  
 QY 3883 agagtgatacccccagagagtgagagcactgtactactactgctgattggaatgagaa 3942  
 Db 2097 agagtgatacccccagagagtgagagcactgtactactactgctgattggaatgagaa 2156

QY 3943 atacaatggaatccacaagaacctgnaattaaatgaatgatagttccaacaaagcaagtg 4002  
 Db 2157 atacaatggaatccacaagaacctgnaattaaatgaatgatagttccaacaaagcaagtg 2216  
 QY 4003 tgttcagcagaataacctcggagagctcatttgtgcttctgtatgtcccaatgtccccc 4062  
 Db 2217 tgttcagcagaataacctcggagagctcatttgtgcttctgtatgtcccaatgtccccc 2276  
 QY 4063 atacctatctctcccaacctggccaatccacagccttgaaggcggaattccacgggggc 4122  
 Db 2277 atacctatctctcccaacctggccaatccacagccttgaaggcggaattccacgggggc 2336  
 QY 4123 agtctcaatcctgacttggagacagctccttgggagatgatattagacattggaacagctcac 4182  
 Db 2337 agtctcaatcctgacttggagacagctccttgggagatgatattagacattggaacagctcac 2396  
 QY 4183 aagtatatcatcgaataaagtatacattcttgatctcagagacaaagtccaatgaaatc 4242  
 Db 2397 aagtatatcatcgaataaagtatacattcttgatctcagagacaaagtccaatgaaatc 2456  
 QY 4243 ctccaagtgaatactactgctcctcatalcccaagaagccaactctggagagctcttttg 4302  
 Db 2457 ctccaagtgaatactactgctcctcatalcccaagaagccaactctggagagctcttttg 2516  
 QY 4303 tttaaacagaaaacattacttlttgaataatggcacaagatcttccattgtcattcaagct 4362  
 Db 2517 tttaaacagaaaacattacttlttgaataatggcacaagatcttccattgtcattcaagct 2576  
 QY 4363 gttgataagtgctgacttgaataatcagaataatccaaatgtcagagatcttgtttatc 4422  
 Db 2577 gttgataagtgctgacttgaataatcagaataatccaaatgtcagagatcttgtttatc 2636  
 QY 4423 cctccacagactccgcagagagacacctgactcgtatgaacgctgtctccttgcctaat 4482  
 Db 2637 cctccacagactccgcagagagacacctgactcgtatgaacgctgtctccttgcctaat 2696  
 QY 4483 attcataatcaacagacacattcctgtgacattacatttaaaatattatgtggaatgata 4542  
 Db 2697 attcataatcaacagacacattcctgtgacattacatttaaaatattatgtggaatgata 2756  
 QY 4543 ggagaactgcagctgttaataagctag 4569  
 Db 2757 ggagaactgcagctgttaataagctag 2783

RESULT 7  
 AAF81927  
 ID AAF81927 standard; cDNA; 2745 BP.  
 XX  
 AC AAF81927;  
 XX  
 DT 13-JUN-2001 (first entry)  
 XX  
 DE Human ICAC-1 nucleotide sequence.  
 XX  
 KW ICAC-1; ICAC-2; asthma; atopic allergy; asthma-related disorder;  
 KW interleukin 9 induced calcium activated chloride channel; IL-9;  
 KW calcium activated chloride channel; anti-allergic; anti-asthmatic;  
 KW anti-inflammatory; immunomodulatory; cystic fibrosis;  
 KW inflammatory bowel disease; autoimmune disease; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX  
 FH Key  
 FT CDS  
 FT 1..2745  
 FT /\*tag= a  
 FT /product= "ICAC-1"  
 FT /note= "IL-9 induced calcium activated chloride channel".  
 XX  
 PN W09944620-A1.  
 XX  
 PD 10-SEP-1999.



Db 1564 tggacaaagcagcctccccaatcctctctggatccccagtgacagaagcagcgtgac 1623  
 QY 3448 ttgtgagtggacaaaacacccaataatgctactcctccaatccagcagcattgtaaggtc 3507  
 Db 1624 ttgtgagtggacaaaacacccaataatgctactcctccaatccagcagcattgtaaggtc 1683  
 QY 3508 ggcacttggaaatcagctctcagcagcagcgcacaaactgaccctgactgctcagctcc 3567  
 Db 1684 ggcacttggaaatcagctctcagcagcagcgcacaaactgaccctgactgctcagctcc 1743  
 QY 3568 cgtgcttccaaatgctacccctgctcccaattacagtgacttccaaacgaaacagacaccc 3627  
 Db 1744 cgtgcttccaaatgctacccctgctcccaattacagtgacttccaaacgaaacagacaccc 1803  
 QY 3628 agcaaatcccccagcctctgtgtatgtatgtacaaatctgcgaaggagcctcccaatt 3687  
 Db 1804 agcaaatcccccagcctctgtgtatgtatgtacaaatctgcgaaggagcctcccaatt 1863  
 QY 3688 ctcaaggcccaatgctacagccctgattgaaatcagtgaaatggaaaacagttactcttgaa 3747  
 Db 1864 ctcaaggcccaatgctacagccctgattgaaatcagtgaaatggaaaacagttactcttgaa 1923  
 QY 3748 ctactggataatggagcagcgtgctgactaaggaatgacggtgtctactcaaggtat 3807  
 Db 1924 ctactggataatggagcagcgtgctgactaaggaatgacggtgtctactcaaggtat 1983  
 QY 3808 ttcaacaactatgacacgaatgctgtagatagacagtgtaaaagtgcggcctctggagagatt 3867  
 Db 1984 ttcaacaactatgacacgaatgctgtagatagacagtgtaaaagtgcggcctctggagagatt 2043  
 QY 3868 aagcagcccaagcagagatgatatcccaagagatggagacgtgtacatccctgctg 3927  
 Db 2044 aagcagcccaagcagagatgatatcccaagagatggagacgtgtacatccctgctg 2103  
 QY 3928 attgagaatgatgaaatcaaatatggaatcccaagacgtgaatataaagatgatt 3987  
 Db 2104 attgagaatgatgaaatcaaatatggaatcccaagacgtgaatataaagatgatt 2163  
 QY 3988 caacacaaagcagatgctgttccagcagaacatcctcggagagctcaattgtgctctgat 4047  
 Db 2164 caacacaaagcagatgctgttccagcagaacatcctcggagagctcaattgtgctctgat 2223  
 QY 4048 gtcccaatgtcccaatcagctgactcttcccaactgtgccaatcacccagcattaaaggc 4107  
 Db 2224 gtcccaatgtcccaatcagctgactcttcccaactgtgccaatcacccagcattaaaggc 2283  
 QY 4108 gaaattcacggggcagcttcatatcatatcgtactggaacagcctcctggggatgattagac 4167  
 Db 2284 gaaattcacggggcagcttcatatcatatcgtactggaacagcctcctggggatgattagac 2343  
 QY 4168 catggaaacagctcaaatatcatatcgaataaagtaacatcttctgatctcagagac 4227  
 Db 2344 catggaaacagctcaaatatcatatcgaataaagtaacatcttctgatctcagagac 2403  
 QY 4228 aagttcaatgaatccttctcagtgaaatcactgctctcctccaaaggagccaactct 4287  
 Db 2404 aagttcaatgaatccttctcagtgaaatcactgctctcctccaaaggagccaactct 2463  
 QY 4288 gaggaagctcttctgtttaaacaggaacatcttcttgaanaatggcacagatctttc 4347  
 Db 2464 gaggaagctcttctgtttaaacaggaacatcttcttgaanaatggcacagatctttc 2523  
 QY 4348 attgctatcagcgtgtgtatgaaggtcgaatcgaataatcagaatctgacga 4407  
 Db 2524 attgctatcagcgtgtgtatgaaggtcgaatcgaataatcagaatctgacga 2583  
 QY 4408 gttatcttcttcttctcctccagacgtccgagagacacacgtctctgtatgaagctc 4467  
 Db 2584 gttatcttcttcttctcctccagacgtccgagagacacacgtctctgtatgaagctc 2643  
 QY 4468 gtcctctgtcctaatatcatatcaacagacacacatcctgtgcatcttcaattaaatt 4527  
 Db 2644 gtcctctgtcctaatatcatatcaacagacacacatcctgtgcatcttcaattaaatt 2703

QY 4528 atgtggaaatggataggagacgtgcagcttcaatagccttag 4569  
 Db 2704 atgtggaaatggataggagacgtgcagcttcaatagccttag 2745  
 RESULT: 8  
 AAH46102  
 ID AAH46102 standard; DNA: 2742 BP.  
 XX  
 AC AAH46102;  
 XX  
 DT 11-SEP-2001 (first entry)  
 XX  
 DE Human CLCA1 coding sequence, SEQ ID NO:4.  
 XX  
 KW Human CLCA1; goblet cell; mouse Gob-5 orthologue; drug screening;  
 KW expression inhibition; antisense therapy; gene therapy;  
 KW chronic obstructive pulmonary disease; bronchial asthma; antiasthmatic;  
 KW ds.  
 XX  
 OS Homo sapiens.  
 XX  
 FH KEY Location/Qualifiers  
 FT CDS 1..2742  
 FT /tag= a  
 FT /partial  
 FT /product= "Human CLCA1"  
 XX /note= "No stop codon given in the specification"  
 PN W0200138530-A1.  
 XX  
 PD 31-MAY-2001.  
 XX  
 PF 22-NOV-2000; 2000MO-JP08232.  
 XX  
 PR 24-NOV-1999; 99JP-0333479.  
 PR 27-APR-2000; 2000JP-0127589.  
 XX  
 PA (TAKE ) TAKEDA CHEM IND LTD.  
 PA  
 PI Nakanishi A, Morita S;  
 XX  
 DR WPI; 2001-355935/37.  
 DR P-PSDB; AAB73716.  
 XX  
 PT New antisense nucleotide, useful for treatment and prevention of  
 PT bronchial asthma, and chronic obstructive pulmonary disease -  
 XX  
 PS Claim 3; Page 82-84; 104pp; Japanese.  
 XX  
 CC The invention relates to an antisense nucleotide targeted to the mouse  
 CC Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart,  
 CC the CLCA1 gene (coding sequence shown in AAH46102). The invention also  
 CC relates to an antibody specific for the Gob-5 protein, medical and  
 CC diagnostic compositions containing the antisense nucleotide or the  
 CC antibody, and methods and kits for screening for compounds which inhibit  
 CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.  
 CC The antisense oligonucleotides and antibody are therefore useful for the  
 CC treatment and prevention of bronchial asthma and chronic obstructive  
 CC pulmonary disease. The present sequence represents the human CLCA1  
 CC gene coding sequence.  
 XX  
 SQ Sequence 2742 BP; 833 A; 616 C; 623 G; 670 T; 0 other;

Query Match 54.0%; Score 2469; DB 22; Length 2742;  
 Best Local Similarity 96.2%; Pred. No. 0;  
 Matches 2578; Conservative 0; Mismatches 5; Indels 96; Gaps 1;

QY 1888 aaggacatgtggccagcagcagcctctgtatctgtcgtgaagctacaggaagcattat 1947  
 Db 160 aaggacatgtggccagcagcagcctctgtatctgtcgtgaagctacaggaagcattat 219

QY	1948	ttcaaaatgtgttcacatttgcattcccgaaacatggaagacaaagcgtcgtactatgttga	2007
Db	220	ttcaaaatgtgttcacatttgcattcccgaaacatggaagacaaagcgtcgtactatgttga	279
QY	2008	ccaaactttagacctacaaatgtctgtagtttctgtgttcgtcgagttcacttccctcaagt	2067
Db	260	ccaaactttagacctacaaatgtctgtagtttctgtgttcgtcgagttcacttccctcaagt	339
QY	2068	aatgataaaccttcaactctgacgtatggtgcgaactgtggaagagaaggttgaaaggaattcac	2127
Db	340	aatgataaaccttcaactctgacgtatggtgcgaactgtggaagagaaggttgaaaggaattcac	399
QY	2128	ctcacctccgtatcttcatctgcaggaaaaaagttagtcgtaaatgtaagaccacaagttaggca	2187
Db	400	ctcacctccgtatcttcatctgcaggaaaaaagttagtcgtaaatgtaagaccacaagttaggca	459
QY	2188	tttttcacataagtggtgtctctacagatgggggagtttagaagataacataatgataag	2247
Db	460	tttttcacataagtggtgtctctacagatgggggagtttagaagataacataatgataag	519
QY	2248	aaatctactatctcaacttggaagaaataacaaagcagtaagatgtctcagcagttacttggt	2307
Db	520	aaatctactatctcaacttggaagaaataacaaagcagtaagatgtctcagcagttacttggt	579
QY	2308	acaaatgtatgtaagaaggtgtcagggagcgcagctgtttacaccaaagaatgtcacattcaat	2367
Db	580	acaaatgtatgtaagaaggtgtcagggagcgcagctgtttacaccaaagaatgtcacattcaat	639
QY	2368	aaagtaacaagaaactctatgaaaaaagatgtagattgttcttccaatcccgccagaagcagag	2427
Db	640	aaagtaacaagaaactctatgaaaaaagatgtagattgttcttccaatcccgccagaagcagag	699
QY	2428	aaggcttctcataatgtyttgtgcacaacatgtytgaattctatagttgaaattctgttacaagaa	2487
Db	700	aaggcttctcataatgtyttgtgcacaacatgtytgaattctatagttgaaattctgttacaagaa	759
QY	2488	aaccacaacaaagaagagcttccaacacagccaataatccaanaatgtcaatctctcgaaagcaca	2547
Db	760	aaccacaacaaagaagagcttccaacacagccaataatccaanaatgtcaatctctcgaaagcaca	819
QY	2548	gaagtgtacccgtgattctgttgaggacttaagaagaaacacacctatgatacaacaagccacca	2607
Db	820	gaagtgtacccgtgattctgttgaggacttaagaagaaacacacctatgatacaacaagccacca	879
QY	2608	aatcccaacttctcatatgtctgcagatgtgacaagaagatgtgtgttagttagttcccttgaca	2667
Db	880	aatcccaacttctcatatgtctgcagatgtgacaagaagatgtgtgttagttagttcccttgaca	939
QY	2668	ttctggaagcagtgagactgtgtaacgcgtctaactcgcgttaataccaagcagcgcagcttttc	2727
Db	940	ttctggaagcagtgagactgtgtaacgcgtctaactcgcgttaataccaagcagcgcagcttttc	999
QY	2728	ctgtctgcagacagctctgcgcggtgtccctgtgttgaggatgtgtacattgtacagttctgccc	2787
Db	1000	ctgtctgcagacagctctgcgcggtgtccctgtgttgaggatgtgtacattgtacagttctgccc	1059
QY	2788	catgtacaaaatgaaatcatacagataaacaagtgtgcagtgtgacagggacaaactctgcacaa	2847
Db	1060	catgtacaaaatgaaatcatacagataaacaagtgtgcagtgtgacagggacaaactctgcacaa	1119
QY	2848	agattactctgcagagaggtttagaaggagagcttcatctgtcagcgggtcttcgatccgcatctt	2907
Db	1120	agattactctgcagagaggtttagaaggagagcttcatctgtcagcgggtcttcgatccgcatctt	1179
QY	2908	actgataatgtgcacaacttgcctgttttccatgacacacagcagttaatggtgagttgga	2967
Db	1180	actgataatgtgcacaacttgcctgttttccatgacacacagcagttaatggtgagttgga	1182
QY	2968	caagaaataatccaataatgtggcctctctgtgcctgtcagcttagtgattagaagaataatoca	3027
Db	1183	caagaaataatccaataatgtggcctctctgtgcctgtcagcttagtgattagaagaataatoca	1203

[illegible]

XX	AAH35019	9	
XX	ID	AAH35019	standard; cDNA; 3109 BP.
XX	AC	AAH35019;	
XX	DT	03-SEP-2001	(first entry)
XX	DE	Human colon cancer antigen encoding cDNA SEQ ID NO:2101.	
XX	RW	Human; colon cancer; colon cancer antigen; diagnosis; detection;	
XX	KW	colorectal carcinoma; ss.	
XX	OS	Homo sapiens.	
XX	PN	WO200122920-A2.	
XX	PD	05-APR-2001.	
XX	PF	28-SEP-2000; 2000WO-US26524.	
XX	PR	29-SEP-1999; 99US-0157137.	
XX	PR	03-NOV-1999; 99US-0163280.	
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	PI	Ruben SM, Barash SC, Birse CE, Rosen CA;	
XX	DR	WPI; 2001-235357/24.	
XX	DR	P-PSDB; MAG75614.	
XX	PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,	
XX	PT	useful for preventing, diagnosing and/or treating colorectal cancers -	
XX	PS	Claim 1; Page 3587-3588; 9803pp; English.	
XX	CC	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon	
XX	CC	cancer-associated nucleic acid molecules (N) and proteins (P), where	
XX	CC	the proteins are collectively known as colon cancer antigens. The colon	
XX	CC	cancer antigens have cyostatic activity and can be used in gene	
XX	CC	therapy and vaccine production. N and P may be used in the prevention,	

Query Match	54.0%;	Score 2467.6;	DB 22;	Length 3109;
Best Local Similarity	96.1%;	Pred. No. 0;		
Matches 2576;	Conservative	2;	Mismatches	6;
			Indels	96;
			Gaps	1;

QY	1880	gacacatgctgacccagcagcattctctgtatctcgtcttgagctacaggaagcgatttatt	1943
Db	49	gcaagagatgacccasgcactctctgtatctcgtttgaagctacaggaagcgatttatt	108
QY	1950	caaaaagttgcccattttgtattcttcbaaacaatgaaagacaagagcgtgactatgtagacc	2009
Db	109	caaaaagttgcccattttgtattcttcbaaacaatgaaagacaagagcgtgactatgtagacc	168
QY	2010	aaaacttgagacactacacaanaatgctgatagttctcgtgtctgagctactcctccagfna	2069
Db	169	aaaacttgagacactacacaanaatgctgatagttctcgtgtctgagctactcctccagfna	228
QY	2070	tgatgaaacctacacatgagcagcagatgtaggccaactgtgagagaagaggtgaaagatccact	2129
Db	229	tgatgaaacctacacatgagcagcagatgtaggccaactgtgagagaagaggtgaaagatccact	288
QY	2130	caccctcgatttcaatttcagagaaanaagtttagcttgataatgnaaccacaggttagagcatt	2189
Db	289	caccctcgatttcaatttcagagaaanaagtttagcttgataatgnaaccacaggttagagcatt	348
QY	2190	tgtccatgagtgaggctcatalcatcagatgtagggagathtatgacagatcaataatgataagaa	2249
Db	349	tgtccatgagtgaggctcatalcatcagatgtagggagathtatgacagatcaataatgataagaa	408
QY	2250	attctactattccaaatgagaagaatacaagcagttaagaatgttcaacgagttattactcgtgac	2309
Db	409	attctactattccaaatgagaagaatacaagcagttaagaatgttcaacgagttattactcgtgac	468
QY	2310	aaatgtagtaagaagaatgagtcagaggagcgagctgcttaccacaaagaatgcacattcaataa	2369
Db	469	aaatgtagtaagaagaatgagtcagaggagcgagctgcttaccacaaagaatgcacattcaataa	528
QY	2370	agtaacaggaactctatgtaaaaagaagatgtaggtttcttccaatcccgccagacggaaga	2429
Db	529	agtwacaggaactctatgtaaaaagaagatgtaggtttcttccaatcccgccagacggaaga	588
QY	2430	ggctctcataatglttgcacacaatgtagtattctctatagttgaattctgttacagaacaaa	2489
Db	589	ggctctcataatglttgcacacaatgtagtattctctatagttgaattctgttacagaacaaa	648
QY	2490	ccacaaacaaagaagcttccaaacagaanaatctaaaatgcaattccggaagcacatggga	2549
Db	649	ccacaaacaaagaagcttccaaacagaanaatctaaaatgcaattccggaagcacatggga	708
QY	2550	agttatccgagatcttgcagagacttlaagaanaacccctctatgacacaacagccacaaa	2609
Db	709	agttatccgagatcttgcagagacttlaagaanaacccctctatgacacaacagccacaaa	768
QY	2610	tcccacctctcatctgctgcagaatgtgacaanaagaattgtgtgtttagtctcttgacaaatc	2669
Db	769	tcccacctctcatctgctgcagaatgtgacaanaagaattgtgtgtttagtctcttgacaaatc	828

QY	2670	tgyaagcavtgyactvgfthaacgcgcctcaatcgtgaactgaatcaagcagccagcttccct	2729
Db	829	tgyaagcavtgyactvgfthaacgcgcctcaatcgtgaactgaatcaagcagccagcttccct	888
QY	2730	gctcacaacagtttgaagcttgggtctgcgtgggttttggaatgfgaatttgaacatgtctgcacca	2789
Db	889	gctcacaacagtttgaagcttgggtctgcgtgggttttggaatgfgaatttgaacatgtctgcacca	948
QY	2790	tgtacaacatgtaacctcaatacagaataaacgvtgacagtgtacagaggaacacacatctgcaccaag	2849
Db	949	tgtacaacatgtaacctcaatacagaataaacgvtgacagtgtacagaggaacacacatctgcaccaag	1008
QY	2850	attacccgcagcagcttccaggaagggagccatctgcagcaggggtcttgatctgcgatttac	2909
Db	1009	attacccgcagcagcttccaggaagggagccatctgcagcaggggtcttgatctgcgatttac	1068
QY	2910	tgtatgtgcaacacatctgcctgttttccatgtacacacagcagttatvgggagtvgcgaca	2969
Db	1069	t-----t-----	1069
QY	2970	agaaaatcccaatttggcctctcttgcgcgcagcttagttagtgaagaaatatccaac	3029
Db	1070	-----gtagttaaggaaataatccaac	1092
QY	3030	tgtatgtatctgtaaatvtgtctgtctgtacagatgtggagaagacaacatataaatgtggtgtt	3089
Db	1093	tgtatgtatctgtaaatvtgtctgtctgtacagatgtggagaagacaacatataaatgtggtgtt	1152
QY	3090	taacgagtcacaacaaagvtgvtgcatatccacacacagtcgcttvgggccctctgcagc	3149
Db	1153	taacgagtcacaacaaagvtgvtgcatatccacacacagtcgcttvgggccctctgcagc	1212
QY	3150	tcaagaactcagagagagctgtgcacaatatgtgcagaagattacagacatattgtctcaagatca	3209
Db	1213	tcaagaactcagagagagctgtgcacaatatgtgcagaagattacagacatattgtctcaagatca	1272
QY	3210	agttcagaacaaatvtgcctcatatgtatctttttggggcccttcatcaaggaaatvtgagatgtt	3269
Db	1273	agttcagaacaaatvtgcctcatatgtatctttttggggcccttcatcaaggaaatvtgagatgtt	1332
QY	3270	cttccacagcttccatccagcttgaaggttaaggatataacccctccagaacagccagttgat	3329
Db	1333	cttccacagcttccatccagcttgaaggttaaggatataacccctccagaacagccagttgat	1392
QY	3330	gaatgycacacagttatctgtgacagcacgcctvgggaaagagacattgtttcttatcacctg	3389
Db	1393	gaatgycacacagttatctgtgacagcacgcctvgggaaagagacattgtttcttatcacctg	1452
QY	3390	gacaaatgcagcttcccaaatctctctcttctgtggatcccaagtvgacagaagcaagtgtgctt	3449
Db	1453	gacaaatgcagcttcccaaatctctctcttctgtggatcccaagtvgacagaagcaagtgtgctt	1512
QY	3450	tgtatgtgacaacaaacccaanaatgtgccttaactcccaatcccaagcatvtgcttaaggttgg	3509
Db	1513	tgtatgtgacaacaaacccaanaatgtgccttaactcccaatcccaagcatvtgcttaaggttgg	1572
QY	3510	caatttggaaatatcagttcttgcagaagccttcacaacaccttgcacccctgactgtlcatcgtccg	3569
Db	1573	caatttggaaatatcagttcttgcagaagccttcacaacaccttgcacccctgactgtlcatcgtccg	1622
QY	3570	tgtgtccaatgttacccttgccttccaatttccagtgatcttccaaaagaaagcaaggaacag	3629
Db	1633	tgtgtccaatgttacccttgccttccaatttccagtgatcttccaaaagaaagcaaggaacag	1692
QY	3630	caaatctcccccagccctctgtgtaattatgtcaaatatctgcagaagagccttccccaattct	3689
Db	1693	caaatctcccccagccctctgtgtaattatgtcaaatatctgcagaagagccttccccaattct	1752
QY	3690	cagggccaggtgttcaacagccctgtatgtaatcatgtgaatvgaaacaaacagttacccctvgaac	3749
Db	1753	cagggccaggtgttcaacagccctgtatgtaatcatgtgaatvgaaacaaacagttacccctvgaac	1812

QY	3750	acggtataatgagcaggtgctctgatactactaagatgacggtgtcttactcaagttat	3809
Db	1813	actgataatgagcaggtgctgatactactaagatgacggtgtcttactcaagttat	1872
QY	3810	cacacattatgacacggaatggttgataacagtgtaaaagtcggggtcttgaggaggttaa	3869
Db	1873	cacacattatgacacggaatggttgataacagtgtaaaagtcggggtcttgaggaggttaa	1932
QY	3870	cgcgcacagcagcagatgataccacagcagatggagacatgataccatccgtctggtat	3929
Db	1933	cgcgcacagcagcagatgataccacagcagatggagacatgataccatccgtctggtat	1992
QY	3930	tgagatgatagaaatacaatctggaatccacccaagacctgaaatbaaagatgatactca	3989
Db	1993	tgagatgatagaaatacaatctggaatccacccaagacctgaaatbaaagatgatactca	2052
QY	3990	acacaaagaagtgtgtttcagcagaataatctctgggaagtcattttggtcttcgatagt	4049
Db	2053	acacaaagaagtgtgtttcagcagaataatctctgggaagtcattttggtcttcgatagt	2112
QY	4050	cccaaatgctcccaatacctgatactcttccacacttgcccaatcaacacagacctgaagcgga	4109
Db	2113	cccaaatgctcccaatacctgatactcttccacacttgcccaatcaacacagacctgaagcgga	2172
QY	4110	aattcacggggcgagctcatatactgacttggaacagctcctcgggagatgatatgacca	4169
Db	2173	aattcacggggcgagctctctcatatactgacttggaacagctcctcgggagatgatatgacca	2232
QY	4170	tggaacagctcacaaagtatatcatctggaataagtaaatatcttctgatactcaagacaa	4229
Db	2233	tggaacagctcacaaagtatatcatctggaataagtaaatatcttctgatactcaagacaa	2292
QY	4230	gttcaatgaactctctcaagtgaaatactactgtctctccatcccaagaaggacacactcga	4289
Db	2293	gttcaatgaactctctcaagtgaaatactactgtctctccatcccaagaaggacacactcga	2352
QY	4290	ggaagtctctttgtttaacacagaaacattacttttgaaatgagacagatctttcat	4349
Db	2353	ggaagtctctttgtttaacacagaaacattacttttgaaatgagacagatctttcat	2412
QY	4350	tgctattcagcgtctgtataaagtcgactctgaaatacagaataatccaactgacagagt	4409
Db	2413	tgctattcagcgtctgtataaagtcgactctgaaatacagaataatccaactgacagagt	2472
QY	4410	atctctgtttatctctccacagactccgcagagagacacactgacctgataagacgtctgc	4469
Db	2473	atctctgtttatctctccacagactccgcagagagacacacactgacctgataagacgtctgc	2532
QY	4470	tctctgtctctaataatcatataccaacagacacactctctgcatcacattttaaaatat	4529
Db	2533	tctctgtctctaataatcatataccaacagacacactctctgcatcacattttaaaatat	2592
QY	4530	gtggaagtgtgataagagaactgcaagctgtcataagcctgag	4569
Db	2593	gtggaagtgtgataagagaactgcaagctgtcataagcctgag	2632
RESULT 10			
AAH46101			
ID	AAH46101 standard; DNA: 2739 BP.		
XX			
AC	AAH46101;		
XX			
DT	11-SEP-2001 (first entry)		
XX			
DE	Mouse Gob-5 coding sequence, SEQ ID NO:3.		
XX			
KM	Mouse Gob-5; murine; goblet cell; human CLCA1 orthologue; drug screening		
KM	expression inhibition; antisense therapy; gene therapy; bronchial asthma		
KM	chronic obstructive pulmonary disease; antiasthmatic; ds.		
XX			
OS	Mus sp.		
XX			



Key Location/Qualifiers  
CDS 1..2739  
FT 1..2739  
FT /\*tag= a  
FT /partial  
FT /product= "Mouse Gob-5"  
FT /note= "No stop codon given in the specification"

MO200138530-AL.  
31-MAY-2001.  
22-NOV-2000; 2000MO-JP08232.  
24-NOV-1999; 99JP-0333479.  
27-APR-2000; 2000JP-0127589.  
(TAKE ) TAKEDA CHEM IND LTD.  
Nakanishi A, Morita S;  
WPI: 2001-355935/37.  
P-PSDB: AAB73715.  
New antisense nucleotide, useful for treatment and prevention of bronchial asthma and chronic obstructive pulmonary disease  
Claim 3; Page 80-82; 104pp; Japanese.  
The invention relates to an antisense nucleotide targetted to the mouse Gob-5 gene (coding sequence shown in AAH46101) or its human counterpart, the CLCA1 gene (coding sequence shown in AAH46102). The invention also relates to an antibody specific for the Gob-5 protein, medical and diagnostic compositions containing the antisense nucleotide or the antibody, and methods and kits for screening for compounds which inhibit the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells. The antisense oligonucleotides and antibody are therefore useful for the treatment and prevention of bronchial asthma and chronic obstructive pulmonary disease. The present sequence represents the mouse Gob-5 gene coding sequence.

Sequence 2739 BP; 784 A; 687 C; 655 G; 613 T; 0 other;

Query Match 33.2%; Score 1517.6; DB 22; Length 2739;  
Best Local Similarity 75.0%; Pred. No. 0;  
Matches 2017; Conservative 0; Mismatches 554; Indels 117; Gaps 5;

1888 aaagacatgtgacccagcatctctgtatctgtcgtgaagctcagcaagcagatttat 1947  
|||||  
160 aaagacatgtgacccagcatctctcattcaccgtgttgaagctcagcaagcagatttac 219  
|||||  
1948 ttcaaaatgttcacatttgatcttgaacatggaagcaagcagcagctatgtgaga 2007  
|||||  
220 ttcaaaaatgttcacatttgatcttcccgagagctggaagcagcagcagctatagcagg 279  
|||||  
2008 ccaaaacttgagacctaacaagaatgtctgtctgtctgtctgtctgtctgtctgtctgt 2067  
|||||  
280 ccaaaacttgagacctaacaagaatgtctgtctgtctgtctgtctgtctgtctgtctgtgc 339  
|||||  
2068 aaatgagacccctacacagatgagcagatgagcagatgagcagatgagcagatgagcag 2127  
|||||  
340 aaatgagacccctacacagatgagcagatgagcagatgagcagatgagcagatgagcag 339  
|||||  
2128 ctcaactcctgattctatctgcaagcaaaagtgtgaatgatgacacacagcagcagcagc 2187  
|||||  
400 cgcagccctgactctttagcaggaagaagcgtgacccagatgagcagcagcagcagcagc 459  
|||||  
2188 ttgttcacatgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2247  
|||||  
460 ttgttcacatgagtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 519  
|||||  
2248 aaattctacttccatggaataaataacagcagatgagcagcagcagcagcagcagcagcag 2307  
|||||

Db 520 aagttcattatccaaagaaaccccaagcagtgagtgcttcagcagcattaccgt 579  
Qy 2308 acaatgtaagaagatgtcagggagcagcgtttacaccaa---aaatgacatc 2364  
|||||  
Db 580 aaaaatcaatctgtcgtgctccagggagcagcagcagcagcagcagcagcagcagcagc 639  
Qy 2365 aataaagtaacagagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2424  
|||||  
Db 640 gacagagtaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 699  
Qy 2425 gaagaagccttctataatgcttgcacaacatgtgattctataatgattctgtacagaa 2484  
|||||  
Db 700 gagaagccttctataatgcttgcacaacatgtgattctgtatgattctgtacagaa 759  
|||||  
Qy 2485 gaaacccacacaaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2544  
|||||  
Db 760 aaaaatcacatcaagaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 819  
Qy 2545 tgggaagtgatccgttgatctgagcagcagcagcagcagcagcagcagcagcagcagc 2604  
|||||  
Db 820 tgggaagtgatccgttgatctgagcagcagcagcagcagcagcagcagcagcagcagc 879  
Qy 2605 ccaatccaccccttctatgctgagcagcagcagcagcagcagcagcagcagcagcagc 2664  
|||||  
Db 880 cctgacacccttctatgctgagcagcagcagcagcagcagcagcagcagcagcagcagc 939  
Qy 2665 aaatctggaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2724  
|||||  
Db 940 aagtcgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 999  
Qy 2725 ttcctgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2784  
|||||  
Db 1000 ttcctgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1059  
Qy 2785 gccactgtaaaaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2844  
|||||  
Db 1060 gctatgtaaaaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1119  
Qy 2845 aaaaatgtaaaaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2904  
|||||  
Db 1120 aagcacttaccacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1179  
Qy 2905 ttaactgatatgtgcaacatctgctgttccatgacacacagcagcagcagcagcagc 2964  
|||||  
Db 1180 tttac----- 1184  
Qy 2965 cgcagaaatccaaatgtgagcctctgctgagcagcagcagcagcagcagcagcagcagc 3024  
|||||  
Db 1185 -----agtgataagaagaat 1203  
Qy 3025 ccaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3084  
|||||  
Db 1204 ccaactgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1263  
Qy 3085 tgccttaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3144  
|||||  
Db 1264 tgccttaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1323  
Qy 3145 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3204  
|||||  
Db 1324 gcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1383  
Qy 3205 gatcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3264  
|||||  
Db 1384 gatcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1443  
Qy 3265 gctgtctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3324  
|||||  
Db 1444 gcatgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1503  
Qy 3325 tggatgaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3384  
|||||  
Db 1504 tggatgaatgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 1563  
|||||



QY	3385	accgagcaaatgcgacctccccaatctctctctgtgatacccaagtcgacagaagaagtc	3444
Db	1564	accgagcaaatgcgacctccccaatctctctctgtgatacccaagtcgacagaagaagtc	1633
QY	3445	ggctcttgatcgcgcaaaaacacacaatatgcctctacatccaaatccagcatctctaa	3504
Db	1624	ggcttatactagacacaaacctaaagtcgcttactctcaagtcaccagcgacgctaa	1683
QY	3505	gttgcaccttggaaatacagctctgcgaagcaagctccaaccttgcacctgtctgcag	3564
Db	1684	gttgcaccttggaaatacagcatccaagcgagctccacaagcatctcaacttacttctcc	1743
QY	3565	tcocgtgtccaaatgtctaccctgcctcccaattacagtgtaacttccaacagcaagac	3624
Db	1744	tcocgtgtcgaagaagtgtaactacgtcctccatctacagtgtaaccggtgtgtgaataagac	1803
QY	3625	accagcaaatccccaagcctctgttagtltatgcacaatatctgcgaagagacctccca	3684
Db	1804	accaggaataatccccagcctctgttaacagtgltatgcacaagcatctgcgaagcctcc	1863
QY	3685	attctcagggcgcaatgtgcacagccctgattggaatcagtgtaatggnaaaacgcttact	3744
Db	1864	attctcagggcgcaatgtgcacagcctctgtatgtaactcgtgtgaataaacaagtaacct	1923
QY	3745	gaactacttgataatgtagcagagtcgtctgatactctaaagatgaaagcgtctactcaag	3804
Db	1924	gaattacttgataatgtagcagagtcgtcgtatgcacaaagatgatgtgtctactcaag	1983
QY	3805	tatttcaaccttctgcacagaatgtgtatagatagatgtataaagtcgcgctctggagga	3864
Db	1984	tttttcaagcttcttgcataaagtgtatagataagcgttataaataatgtgccttggagga	2043
QY	3865	gttaacgacgacaaacgagagatgataccccaagcagagtcgcacttacaactctgc	3924
Db	2044	gtcaacttcagacagacagagagacgaacacctccgaagaagaagccatgtactatagatgc	2103
QY	3925	tgtatctggaatgtatgataacaaatggaaatcccaagaacctgtaaataaagatgat	3984
Db	2104	tgtatctgagatgtgtgaagaatgaaatgaaaccccaacgctcgtgaactagt-----at	2157
QY	3985	gttcaacaacaagcagagtcgttcttcacagaaaactctctggagcgtactattgtgctctc	4044
Db	2158	gttcaacaacaagcagcgtgtcttcacagaaaactctccggggagtcgttctgtgcacc	2217
QY	4045	gatgtccc--caaatgtctccatactgtactcttcccaacctgcgcaaatcacccagctg	4101
Db	2218	aatgtccocgagagatctccatctctgactcttcccaacctgtccaataactgactg	2277
QY	4102	aagcggaataatcacggtgggacgtctcatatctgacttggacagctcctcgggagat	4161
Db	2278	aagcgacgacatccaagggcgagaaacctcggtgaaatctgacgtctgcgcctcggtgagtac	2337
QY	4162	tatgacaatggaaacgctcaagaatgatatacatctcgatagaagcaagatcttgatctc	4221
Db	2338	tagacaacaagggagagcttccaactatcatcatccgaatgagacaagcatcgttgatctc	2397
QY	4222	agagacaagtcaatgaaatctcttcacagtgaatactactgtctcatcccaagaagacc	4281
Db	2398	agggacaacctatcaaacactcatctccaagtgaaacaatacgcgtctatctcccaagagcc	2457
QY	4282	aacctgaagaagtccttgtttaaacaagaataactactttgaaatggacaagat	4341
Db	2458	agctcgtgagaaactcttgatcttgacatcggtggaggaacaactttggaaatggacagat	2517
QY	4342	cttttactgtcatctacgctgtctgtgataagtcogactgaaatccgaataatccaacat	4401
Db	2518	attctcaatgtcatccagcgtgtgagtaagtcocaaatctgaaatccgaataatccaacat	2577
QY	4402	gcaagagatccttgttattatctctccaagactccgcagagacaacctagtctgtatgaa	4461
Db	2578	gcaagagagtcctgtgttcatctcccgctgc-----gagcgcgcattctccgaagac	2628

Accession	Sequence	Position
Oy	acgtgtgcttcttctaataatcatcaacagacattccgtgcaatcaattta	4521
Oy		4522
Oy		4523
Oy		4524
Oy		4525
Oy		4526
Oy		4527
Oy		4528
Oy		4529
Oy		4530
Oy		4531
Oy		4532
Oy		4533
Oy		4534
Oy		4535
Oy		4536
Oy		4537
Oy		4538
Oy		4539
Oy		4540
Oy		4541
Oy		4542
Oy		4543
Oy		4544
Oy		4545
Oy		4546
Oy		4547
Oy		4548
Oy		4549
Oy		4550
Oy		4551
Oy		4552
Oy		4553
Oy		4554
Oy		4555
Oy		4556
Oy		4557
Oy		4558
Oy		4559
Oy		4560
Oy		4561
Oy		4562
Oy		4563
Oy		4564
Oy		4565
Oy		4566
Oy		4567
Oy		4568
Oy		4569
Oy		4570
Oy		4571
Oy		4572
Oy		4573
Oy		4574
Oy		4575
Oy		4576
Oy		4577
Oy		4578
Oy		4579
Oy		4580
Oy		4581
Oy		4582
Oy		4583
Oy		4584
Oy		4585
Oy		4586
Oy		4587
Oy		4588
Oy		4589
Oy		4590
Oy		4591
Oy		4592
Oy		4593
Oy		4594
Oy		4595
Oy		4596
Oy		4597
Oy		4598
Oy		4599
Oy		4600
Oy		4601
Oy		4602
Oy		4603
Oy		4604
Oy		4605
Oy		4606
Oy		4607
Oy		4608
Oy		4609
Oy		4610
Oy		4611
Oy		4612
Oy		4613
Oy		4614
Oy		4615
Oy		4616
Oy		4617
Oy		4618
Oy		4619
Oy		4620
Oy		4621
Oy		4622
Oy		4623
Oy		4624
Oy		4625
Oy		4626
Oy		4627
Oy		4628
Oy		4629
Oy		4630
Oy		4631
Oy		4632
Oy		4633
Oy		4634
Oy		4635
Oy		4636
Oy		4637
Oy		4638
Oy		4639
Oy		4640
Oy		4641
Oy		4642
Oy		4643
Oy		4644
Oy		4645
Oy		4646

RESULT	11
AAH46120	
ID	AAH46120 standard; cDNA; 2843 BP

AC AAH46120;

DT 11-SEP-2001 (first entry)

Mouse Gob-5 cDNA, SEQ ID NO:22.

KW Mouse Gob-5; murine; human CLCA1 orthologue; drug screening;

chronic obstructive pulmonary disease; antiasthmatic; ss.

Mus sp.

FH	Key	Location/Qualifiers
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FF

**XX**

XX

XX

XX

PR 27-APR-2000; 2000JP-0127589

PA (TAKE ) TAKEDA CHEM IND LTD.

PI Nakanishi A, Morita S;

DR WPI; 2001-355935/37.

XX  
XX

PT bronchial asthma and chronic obstructive pulmonary disease

PS Example 1; Page 89-91; 104pp; Japanese.

CC The invention relates to an antisense nucleotide targetted to the mouse

the CLCA1 gene (coding sequence shown in AAH46102). The invention

CC diagnostic compositions containing the antisense nucleotide or the

CC the protein. Gob-5 and CLCA1 are proteins expressed by goblet cells.

## CC treatment and prevention of bronchial asthma and chronic obstructive

XX

Sequence 2843 BP; 815 A; 705 C; 680 G; 643 T; 0 other;

Query Match 33.28; Score 1517.6; DB 22; Length 2843;

Matches 2017; Conservative 0; Mismatches 554; Indels 117; Gaps 5;

QY 1888 aagacatcgtgaccagcatctctgtatctgtgaagctacaggaagcgatttat 1947  
 |||||  
 Db 174 aagacatcgtgactcagcctctccatactgttttaagctacaagaaaaaatttac 233  
 |||||

QY	1948	ttcaaaaatgttgcattttgatttcttcttgaaacatggaaagacaagaagctgcacatgtgtgaa	2007
Db	234	ttcaaaaatgttgcattttgatttcttcttgaaacatggaaagacaagaagctgcacatgtgtgaa	293
QY	2008	ccaaactctggaacctcaaaaatgtgcagatgtgtctgttcgtgcagtgctactctccagat	2067
Db	294	ccaaactctggaacctctcaaaaagcgtgatgtctctgttatcaaacaccagccctctgaac	353
QY	2068	aatgatgaaccttaacctgaactgacagatggtggaactgttgagagaaaggtgtgaaagatccac	2127
Db	354	aatgatgagcccttaacctgaactgaataatagagcatgttgagagaaaggtgtgaaagatccac	413
QY	2128	ctcactctgatttccattgttcaggaagaaaagttagctgcgaataatgagccacaaagttaggca	2187
Db	414	ctgcactctgactcttcttagagagaagaagctgtacccaatgtatgggcccacaaagacagacc	473
QY	2188	tttctgcatagtgtggtctctactacgatctggagagatatttgaaagatatacatatgatag	2247
Db	474	tttgcctcatgtgtggtctactctccgatctggagaggtgtttaaagataatacaaaagacagag	533
QY	2248	aaattctactatccaaatggagaataatacaagacagtaagatgttcttcagcagttacttgt	2307
Db	534	aagtctactatctccaaagaaaaccccacagacagtgaggtgtctcagcagccattacggt	593
QY	2308	acaaatgtatgtaagaaggtgtctcagagagcagctgtgttaaccaa---aagatgcacatlc	2364
Db	594	aaaaatacagttctgcgtgcgcagagaggaagctgtgtatactcaacaggaaggtgttaac	653
QY	2365	aataaagttaacaagagaccttatgaaagaagtgtgagttgtcttccaaatcccgcacagcg	2424
Db	654	gacagagtaaacggtgacctgtataagacaaattgtgtattgttccagatccacacccaacac	713
QY	2425	gagaaagcttctataatgtttgtgcacacaatgttgatctcatalgttgaattctgttaagaa	2484
Db	714	gagaaagcttctccatcatgttttaacaaaatatcaattctgtgttgaattctgttaagaa	773
QY	2485	caaaatcccaacaagaagagctccaacaagcaaatccaatctgaatctccgaagacaca	2544
Db	774	aaaaatacacaatccaagaagaccccaatatgacaacaaacacagatctcgaatccgaagacag	833
QY	2545	tggagaatgtgacgtgtgattcttgagagctttaagaaacacaccctcatgacaacaagaca	2604
Db	834	tggagaatgtatccacagaaatcttgagagactccaagcaaacacccatccatgacacgccagaca	893
QY	2605	ccaaatcccaacctctcaatttgcctgcagatgtgacaaagaagattgtgtttagtctcttac	2664
Db	894	ccctgaccacacacttctcaactgcctgcagaaattgacaaagaattgtgtcttagtctctgat	953
QY	2665	aaatcttgaaagcatgtgcgactgtgtaacgcgctccatccgatctgaatcaagcagacagctt	2784
Db	954	aagctcgcggagcatgtgcgaacagatgtatctgtcttaaacccgaatatcaatcgaacgcgctgt	1013
QY	2725	ttctcgtgcgcgaagaagtttgagctggtgggtctctcgtgtgttggaatgtgaacttgaacgtct	2784
Db	1014	ttctcgtgcgcgaagatctgtgagagcaggatctccgtgtgtctgggaatgtgaacttgaacgtct	1073
QY	2785	ggccaatgttacaanaatgtgaacctacatcacgataaacaagctgcagtgcaagggagacacactcc	2844
Db	1074	gctcatgttacaagaagcgaacctcaaaacagtttaaacagtggtgcctgacaagagatctgctgac	1133
QY	2845	aaaaagatatacctgcagcagctcttaagagaggaagtcacatctgcagcgggtcttcgatcggca	2904
Db	1134	aagaactataccacagatatactctgcaagagacatataatgtctctgcgctcttcggacgacga	1193
QY	2905	tttactgatataatgtggaacaaattgcctgttttccatagcaacacagcttatgtggaggtg	2964
Db	1194	tttctac-----	1198
QY	2965	gcacagaagaataatccaatgttgctctctctgcgtcgtcagctttagtattagaagaataat	3024
Db	1199	-----agtgataaagaagaagatcat	1217
QY	3025	ccaactgtatgatactgaattgtctgcctgcagcagatgtgggaaagaaacaaactataagtgtg	3084

Db	1218	ccaactgaatgactctgaaatttgctgagctgcaacggaatgaggagaaacaacccattagcagc	1277
QY	3085	tgctctlaacgaggtgctcaaaacaagttggtgcacatcatccacagtlcgcttggggccctc	3144
Db	1278	tgctttgacgcgtggaagcagagcgaggccatcatccatacagctggtccctgggacgcgt	1337
QY	3145	gacagctcaagaatacagagagcgtgccaaaatgacagagaggtttacagacataatgctca	3204
Db	1338	gcccataaagagctcttgacgagctgtccaaaatbaccagagagcgcttcgaacatactctcg	1397
QY	3205	gataaagtccagaaacaaatggccctattgagcttctggggccctttccaaagaaatgta	3264
Db	1386	gatacaggtctcgaacaaatggtccttgatgagcttctgcagacactccctccagaaatgctg	1457
QY	3265	gcttctctccagcgtccatccacgagcttgagagtaagaaatlaacccctccagaaacgagca	3324
Db	1458	gcgagctgcctcagacactccatccacgctggagagagaggggttaactctccagataaccaa	1517
QY	3325	tgagatgaatgagcagatgatacgttggaacgacccgttggaagaagacatttgttctatc	3384
Db	1518	tgagatgaatgagcgtctcagtgatcgtggaacgctggttggaagaagacatttgttctatc	1577
QY	3385	accctggcaatgacagcctccccaatccctctcttgagatcccaatgagacagaagaagt	3444
Db	1578	accctggcaagaagcactccctcaataatattatctggatcccaagggagtggaacaatat	1637
QY	3445	gagcttgtagtgagcaaaaacacccaataatgagctacatcccaatcccgagctatga	3504
Db	1638	ggtttatatacagacaaacacctaagtggtgctacatcccaatcccgagcagcgctaa	1697
QY	3505	gttgagcaattggaataatcagctgtcgaagcaagctcaacaaacttgacctgagctgacg	3564
Db	1698	gttgagcttttggaataatcagcatccaacgagctcagacactccaacttgagctgacac	1757
QY	3565	tcccggtgctlccaatggttaccgctccatccaatgaagtgactccaagaacgaagaagac	3624
Db	1758	tcccggtgagcaagtgctataacagctccctatbaccagtgagcccggtatgagataagaac	1817
QY	3625	accgagcaaatcccccaacccctctgtagttatgacaaatattcgcgaagagcctcccca	3684
Db	1818	acagaggaattccccagcctctgtaacagtggtatgcaagcattcgcgaagagcctcgct	1877
QY	3685	attctcagggccagtgctacagacccctgattgtaacagtgaaatgynaacaaagattactctg	3744
Db	1878	attctcagggccagcgtcagacagccttgattgtaacatctgtgaatgynaacaaagattactctg	1937
QY	3745	gaactactgtaataatgagcaggtgctgtatgcttactaagaatgacggtgtctactcaag	3804
Db	1938	gaactactgtaataatgagcaggtgctgtatgcttactaagaatgacggtgtctactcaag	1997
QY	3805	tattcccaactttagcagaaatgtaataacagtgtaaaagtgcgggctctggagga	3864
Db	1998	tattcccaactttagcagaaatgtaataacagtgtaaaagtgcgggctctggagga	2057
QY	3865	gttaacgcagccagacgagagtgataaccaccaagtgagacatglaataactgyc	3924
Db	2058	gtcaacttcagacagacagagagcagacacttcggaagaacagagccatgtlaactagatgyc	2117
QY	3925	tgagatgagaatgataatgataacaaatggaaatccaacgaagcgtgaatatgaagaatgac	3984
Db	2118	tgagatgagaatgataatgagaatgaagaatgaacacacacgctctcgaataactatgtl-----at	2171
QY	3985	gttaacaacaagcagagtggttttcagacagaatactctggagaggttcaattgttggtctc	4044
Db	2172	gttcaaaagacagcagcgtgtcttcagacagaacatcttcaaggggatcgtctgtggtccaac	2231
QY	4045	gatactc-----caaatgctccacatactgtctctccacgtgagcaatatcagacatg	4101
Db	2232	aatgtcccgagcagctcccatctctgacctcttccacccctgtfcaataactcagactg	2291
QY	4102	aagcggaattcaacggggtcagctcattcaatattcgtacttgagacagcctctggagatg	4161



OY	2545	tgggaaagtgcacgcttgattcttgagactttaaagaaacaaccccttatggcaacagcgca	2604
Db	827	tgggaaagtgcacgagaaacttgagagacttcaagcaaaacaaccccttgacgcccagcgca	886
OY	2605	ccaatccaccccttcattctgtctgcagatctgagcaaaagaaattgtgtttaagtccttgac	2664
Db	887	cttgacaccaccttctcaactgcgtgcgaatttgagcaaaagaaattgtgttgcttgattgat	946
OY	2665	aaatctggaagcatgctgcagactgtgtaacgcctcaatcgactgtaatcaagagcgacatt	2724
Db	947	aagtcgcggagcatgctgcgaacgatgatgcgtcttaacccggaatgaatgaagcaagccgctt	1006
OY	2725	ttccctgcgcgagaaagtctgagcggtggtcccttggttggtgaattgtgaattgacagtct	2784
Db	1007	ttccctgcgcgagactgtgagcgagggagctccctgggttcggatgtgtgcctttgacagtctt	1066
OY	2785	gcccattgacaanaatgtgaacccatatacagataaacaagctgcagctgagacagacacatccg	2844
Db	1067	gctcatgtacaanaagcggaaatccaacagttaaacaagtggtgcctgcagagagatctgtctgc	1126
OY	2845	aaaagataacctgcagcagactctcaagagggagctccatctgcagcggtcttcgatcggca	2904
Db	1127	aagcaacttaccacaagatattcgaagagggagacatctatagtctctgcgtccctgcagacga	1186
OY	2905	tttactgataatgtgcaacaatttgcctgtttccatagacaacaagagttatggggagt	2966
Db	1187	tttac-----	1191
OY	2965	cgacaagaanaatccaatcttggcctctctgcgtccctgcagcttgattgatagaanaatat	3024
Db	1192	-----agtataaagaagaagatat	1210
OY	3025	ccaactgatactggaatctggaattgtctgtctgtaagaagatctgagaaacaacataataagttg	3084
Db	1211	ccaactgatactggaattgtctgtctgtaaccgtaaggagagacaacaacataagtcagacg	1270
OY	3085	tgctttaacagagtgcaacaanaagtgtgtgcataatccaacagtcgctttggggccctc	3144
Db	1271	tgctttgcctctgtgtaagagagagcggtggccatatccatacgaattggtcccttgagccgctt	1330
OY	3145	gcaagctcaagaagcatagagggcgctgccaaaatgacaggggtttacaagacatagtcttca	3204
Db	1331	gcccgttaagaagccttgtagcagcgctgtccaaaatgacagggggcctgcagacataactcttcg	1390
OY	3205	gatacaagttcgaacaaatgtagcctcatatgatgcttttgggggccccttcacaaagaaatgta	3264
Db	1391	gatacaggttcgaacaaatgltctgtgtgatgcttttcgacgactccctccaggaatgcg	1450
OY	3265	gctgtctctcaagcgctccatccacagcttgaaagtgaaggaattaaacctccagaagaacgca	3324
Db	1451	gcgatcgtctcagcactccatccacagcttgaaagagcaggtggaattatccagaataaaca	1510
OY	3325	tggaatgaaatgacagatgacgcgtgagacagaccgttgagaaagacaacttgctttttatc	3384
Db	1511	tggaatgaaatgctcagtgatcgcgtgagacagctctgggtgaggaagacaacttgctttttatc	1570
OY	3385	aacttgacaacatgtagcgtctccccaactctctctctctgttgatcccaagtggacaagaagaagt	3444
Db	1571	aacttgacaacagcatctctccacaacataattatcttggtatcccgagcgagttgacaanaat	1630
OY	3445	ggccttctgagtgtgagcaaaaacaacccaatgtgcctactcccaatccagagcatgtctag	3504
Db	1631	ggtttattactagacaacaacccaatgaagtgtgctactcccaagtcccaagcgacggtcttag	1690
OY	3505	gttgagcaattgaaatacagctctgcgaagcaggtgacaaaccttgacccttgactgtacg	3566
Db	1691	gttgagcttttgaaatacagcatctcaagcgagctccagacttcaaccttgactgtacgc	1750
OY	3565	tcgccgtgtgtccaaatgtataccctctggtccctcccaattagttacttccaaacggaacaagac	3624
Db	1751	tcgccgtgtgagaaagtgtctacacgtgcctcccatatacagtcaccccggttagttgataagaac	1810
OY	3625	accagcaaatccccaagcctctctgtgtattatgcaaatattctgcgaaggaagcctcccca	3684

Db	1811	acagggaattcccccagcctctgtataacagtgatgtcaagcatctgcgcaaggagcctcgct	1870
QY	3685	atctccagggcagtgctcacagcccttgatgtaatcagtgatgtgaaaaacagttactcttg	3744
Db	1871	attctcaaggccagcgtcacagccttgatgtaatctgtgtaatgtgaaaaacagtaacctg	1930
QY	3745	gaactcactgataatgtgagcagtgctctgatagtcactaaagatgagcgggtgctactcaag	3804
Db	1931	gaattactgtgataacggagcaggtgcccagatgcacacaaagatgattgtgtactcaag	1990
QY	3805	tatttcaacaatttgacacgaaatgtgtagaatacagtgtaaaagtgcggctctgtggagga	3864
Db	1991	ttttttacagcttttgatgtcaaatgtgtagatacagcggttaaaatatggtcctgtggagga	2050
QY	3865	gttaacgacgacagacgagatgtatatacccaagcagatgtggacactgtatactactg	3924
Db	2051	gtcacttcaagcagacagagagacagcactcgcgaagaacagagcatgtatactatgtgc	2110
QY	3925	tgtgattgagaaatgtatgaataacatgtgaatcccaagccctggaattataaagatgat	3984
Db	2111	tgtgattgagatgtgtgtgaagatgaatgaaccccaacgcttccttgaaactagt-----at	2164
QY	3985	gttcaacaacagcagatgtgttttcagacagaaactccctggggagctcatgttgtgctct	4044
Db	2165	gttcaagaacagcagcgtgtgtcttcagcaggaacatcttccaggggagtcgtttgtgcca	2224
QY	4045	gatytc---caaatgtcccatcactgtactcttcccaactgtgccaatacagcactg	4101
Db	2225	aatgtcccccagcagctcccatctcactgtactcttcccaactgttcaactcactgtactg	2284
QY	4102	aagcgcggaattcagcggggcagtgctcalttaactctgactgtgacagctctctgtggatgat	4161
Db	2285	aaggccagcagctccaaaggcagaaactctgtgtgaactctgacgttggaagcgtctctgtggatgac	2344
QY	4162	tatatacctgtgaaagcgtccaaagtatatacttcgtgaataagatgatacttctgatactc	4221
Db	2345	tacgacccagggagagcttcccaactacatccagatgagacccagatagctgtgatactc	2404
QY	4222	agagacaagtatcagatgaactcttcttcaagtgaatactactgtctcatcccaaggaagcc	4281
Db	2405	agggaaccaactcaacactcactccaaagtgaacactaacctacgcttatccccaaagaagcc	2464
QY	4282	aactctgaggaagctcttttgtttaaacagaaacattactcttggaaaaatgtgcagat	4341
Db	2465	agctctgaggaacacttctgttgatttgtaactgtggaggaacacacttctgtgaattgtccacagat	2524
QY	4342	ctttcaattgtatctagcgtctgttgaagaagtcgtgactctgaatctgaaatctgaatactcaact	4401
Db	2525	aactcattctgataccagcgtctgtgataaagatccaaactcgtaaatcagaatctcccaact	2584
QY	4402	gcaagcagatcttctgttatcttccatcccaagactccgcagagaacacactagtcgtatgaa	4461
Db	2585	gcaagcgggtgtgtgttcatcccccgcgcag-----gagcgcgcacttcccggaagc	2635
QY	4462	acgtctgtcctctgtcctaataatcatatacaacagcacatctctgtgacatacttta	4521
Db	2636	tcaactccccccttgctctgatactagatgaataacagacacttctgtgatacagtgctg	2695
QY	4522	aaaattatgtgaagtgtgataagagaactgcacttcaataagcctag	4569
Db	2696	aagataatgtggaagtgtgcataagggaatgtcagtgagacactgattgtg	2743
RESULT 13			
AA265095			
ID AA265095 standard, cDNA, 3265 BP.			
AC AA265095;			
XX			
DT 05-APR-2000 (first entry)			
DE Membrane-bound protein PRO1124 encoding cDNA.			

XX Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;  
KW pharmaceutical; receptor immunoadhesin; gene mapping; ss.  
XX  
OS Homo sapiens.  
XX MO9963088-A2.  
XX  
PD 09-DEC-1999.  
XX  
PF 02-JUN-1999; 99WO-US12252.  
XX  
PR 02-JUN-1998; 98US-0087607.  
PR 02-JUN-1998; 98US-0087609.  
PR 02-JUN-1998; 98US-0087759.  
PR 03-JUN-1998; 98US-0087827.  
PR 04-JUN-1998; 98US-0088021.  
PR 04-JUN-1998; 98US-0088025.  
PR 04-JUN-1998; 98US-0088028.  
PR 04-JUN-1998; 98US-0088029.  
PR 04-JUN-1998; 98US-0088030.  
PR 04-JUN-1998; 98US-0088033.  
PR 04-JUN-1998; 98US-0088326.  
PR 05-JUN-1998; 98US-0088167.  
PR 05-JUN-1998; 98US-0088202.  
PR 05-JUN-1998; 98US-0088212.  
PR 05-JUN-1998; 98US-0088217.  
PR 09-JUN-1998; 98US-0088655.  
PR 10-JUN-1998; 98US-0088722.  
PR 10-JUN-1998; 98US-0088730.  
PR 10-JUN-1998; 98US-0088734.  
PR 10-JUN-1998; 98US-0088738.  
PR 10-JUN-1998; 98US-0088740.  
PR 10-JUN-1998; 98US-0088741.  
PR 10-JUN-1998; 98US-0088742.  
PR 10-JUN-1998; 98US-0088810.  
PR 10-JUN-1998; 98US-0088811.  
PR 10-JUN-1998; 98US-0088824.  
PR 10-JUN-1998; 98US-0088825.  
PR 10-JUN-1998; 98US-0088826.  
PR 11-JUN-1998; 98US-0088838.  
PR 11-JUN-1998; 98US-0088861.  
PR 11-JUN-1998; 98US-0088863.  
PR 11-JUN-1998; 98US-0088876.  
PR 12-JUN-1998; 98US-0089090.  
PR 12-JUN-1998; 98US-0089105.  
PR 16-JUN-1998; 98US-0089440.  
PR 16-JUN-1998; 98US-0089512.  
PR 17-JUN-1998; 98US-0089514.  
PR 17-JUN-1998; 98US-0089532.  
PR 17-JUN-1998; 98US-0089538.  
PR 17-JUN-1998; 98US-0089598.  
PR 17-JUN-1998; 98US-0089599.  
PR 17-JUN-1998; 98US-0089600.  
PR 17-JUN-1998; 98US-0089653.  
PR 18-JUN-1998; 98US-0089801.  
PR 18-JUN-1998; 98US-0089907.  
PR 19-JUN-1998; 98US-0089947.  
PR 19-JUN-1998; 98US-0089948.  
PR 19-JUN-1998; 98US-0089952.  
PR 22-JUN-1998; 98US-0090246.  
PR 22-JUN-1998; 98US-0090252.  
PR 22-JUN-1998; 98US-0090254.  
PR 23-JUN-1998; 98US-0090343.  
PR 23-JUN-1998; 98US-0090355.  
PR 24-JUN-1998; 98US-0090429.  
PR 24-JUN-1998; 98US-0090431.  
PR 24-JUN-1998; 98US-0090435.  
PR 24-JUN-1998; 98US-0090444.  
PR 24-JUN-1998; 98US-0090445.  
PR 24-JUN-1998; 98US-0090461.  
PR 24-JUN-1998; 98US-0090472.

PR 24-JUN-1998; 98US-0090535.  
PR 24-JUN-1998; 98US-0090538.  
PR 24-JUN-1998; 98US-0090540.  
PR 24-JUN-1998; 98US-0090557.  
PR 25-JUN-1998; 98US-0090676.  
PR 25-JUN-1998; 98US-0090678.  
PR 25-JUN-1998; 98US-0090688.  
PR 25-JUN-1998; 98US-0090690.  
PR 25-JUN-1998; 98US-0090691.  
PR 25-JUN-1998; 98US-0090694.  
PR 25-JUN-1998; 98US-0090695.  
PR 26-JUN-1998; 98US-0090696.  
PR 26-JUN-1998; 98US-0090862.  
PR 01-JUL-1998; 98US-0091358.  
PR 01-JUL-1998; 98US-0091360.  
PR 01-JUL-1998; 98US-0091544.  
PR 02-JUL-1998; 98US-0091478.  
PR 02-JUL-1998; 98US-0091486.  
PR 02-JUL-1998; 98US-0091519.  
PR 02-JUL-1998; 98US-0091626.  
PR 02-JUL-1998; 98US-0091628.  
PR 02-JUL-1998; 98US-0091633.  
PR 02-JUL-1998; 98US-0091646.  
PR 07-JUL-1998; 98US-0091673.  
PR 07-JUL-1998; 98US-0091978.  
PR 09-JUL-1998; 98US-0091982.  
PR 10-JUL-1998; 98US-0092182.  
PR 20-JUL-1998; 98US-0092472.  
PR 30-JUL-1998; 98US-0093339.  
PR 04-AUG-1998; 98US-0094651.  
PR 04-AUG-1998; 98US-0095282.  
PR 04-AUG-1998; 98US-0095285.  
PR 04-AUG-1998; 98US-0095301.  
PR 04-AUG-1998; 98US-0095302.  
PR 04-AUG-1998; 98US-0095318.  
PR 04-AUG-1998; 98US-0095321.  
PR 04-AUG-1998; 98US-0095325.  
PR 10-AUG-1998; 98US-0095916.  
PR 10-AUG-1998; 98US-0095929.  
PR 10-AUG-1998; 98US-0096012.  
PR 11-AUG-1998; 98US-0096143.  
PR 11-AUG-1998; 98US-0096146.  
PR 12-AUG-1998; 98US-0096329.  
PR 17-AUG-1998; 98US-0096757.  
PR 17-AUG-1998; 98US-0096766.  
PR 17-AUG-1998; 98US-0096768.  
PR 17-AUG-1998; 98US-0096773.  
PR 17-AUG-1998; 98US-0096791.  
PR 17-AUG-1998; 98US-0096867.  
PR 17-AUG-1998; 98US-0096891.  
PR 17-AUG-1998; 98US-0096894.  
PR 17-AUG-1998; 98US-0096895.  
PR 17-AUG-1998; 98US-0096897.  
PR 18-AUG-1998; 98US-0096949.  
PR 18-AUG-1998; 98US-0096950.  
PR 18-AUG-1998; 98US-0096959.  
PR 18-AUG-1998; 98US-0096993.  
PR 18-AUG-1998; 98US-0096960.  
PR 18-AUG-1998; 98US-0097022.  
PR 19-AUG-1998; 98US-0097141.  
PR 20-AUG-1998; 98US-0097218.  
PR 24-AUG-1998; 98US-0097661.  
PR 26-AUG-1998; 98US-0097951.  
PR 26-AUG-1998; 98US-0097952.  
PR 26-AUG-1998; 98US-0097954.  
PR 26-AUG-1998; 98US-0097955.  
PR 26-AUG-1998; 98US-0097971.  
PR 26-AUG-1998; 98US-0097974.  
PR 26-AUG-1998; 98US-0097978.  
PR 26-AUG-1998; 98US-0097979.  
PR 26-AUG-1998; 98US-0097986.  
PR 26-AUG-1998; 98US-0098014.  
PR 31-AUG-1998; 98US-0098525.

PR 16-SEP-1998; 980US-0100634.  
PR 12-JAN-1999; 990US-0115565.  
PA (GETH ) GENENTECH INC.  
XX  
XX  
PI Baker K, Chen J, Goddard A, Gurney AL, Smith V, Matanabe CK;  
PI Wood WJ, Yuan J;  
XX WPI: 2000-072883/06.  
DR P-PSDB: AAV66749.  
XX  
XX  
PT Membrane-bound proteins and related nucleotide sequences -  
PS  
XX  
XX Claim 2; Fig 273; 822pp; English.  
XX  
CC The invention provides membrane-bound PRO polypeptides and  
CC polynucleotides encoding them. The PRO sequences of the invention were  
CC identified based on extracellular domain homology screening. The PRO  
CC sequences have homology with proteins including LDL receptors, TIE  
CC ligands and various enzymes. The membrane-bound proteins and receptor  
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor  
CC immunoadhesins, for instance, can be used as therapeutic agents to block  
CC receptor-ligand interactions. The membrane-bound proteins can also be  
CC employed for screening of potential peptide or small molecule inhibitors  
CC of the relevant receptor/ligand interaction. The PRO encoding sequences  
CC are useful as hybridization probes, in chromosome and gene mapping and in  
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences  
CC will also be useful for the preparation of PRO polypeptides, especially  
CC by recombinant techniques.  
XX  
XX Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;

Query Match	24.7%	Score 1129.8;	DB 21;	Length 3265;
Best Local Similarity	67.1%;	Prod. No. 2e-300;		
Matches 1749;	Conservative	0;	Mismatches 747;	Indels 111; Gaps 6;
OY 1889	aggacatggtgcccaggcactctcgatctcgtcattcgtcttgaagctcaggagaagcgatttatt	1948		
DB 182	aggatattggtgactcagcttccactcgttctcgttcttggaaagccagaaaaaagatttttt	241		
OY 1949	tcaaaaatgtytccatttggatttccttgaacaatggagaacaaggctgtaactgtgtgagac	2008		
DB 242	tcaaaaatgtatctatattatatttcctgtggaatttggagaagaaatcctcagtaaaaaagc	301		
OY 2009	caaaacttgagagcctcacaaaaatgctgtatgtctcgtgtgtgcctgactcctccaagta	2068		
DB 302	caaaaactgtaaaaacataacaatcgtctgtatgttatgtgttcaccacctcacctcccaagta	361		
OY 2069	atgatgaaccctcacactgagcagatcggggaactgtggagaagaagggttaaaagatccacc	2128		
DB 362	gagatgaaccatcacacccaagcgacttccacagaattcgggagaagaaggcggaatacattcact	421		
OY 2129	tcaactcgtatttcatcttgcaaggaaaaaagttagctcgtaattatgaccacacaagtaaggcat	2188		
DB 422	tcaacctcgactcttacttcttggaaaaaaaacaatatgacatattgaccacagccaactgt	481		
OY 2189	tgttcacatgagtggggtcctcatctacgaatggggagatatttggacgagtaacataatgtatgaga	2248		
DB 482	tgttcacatgagtggggtcctcaccccggttggggagtttttgaatgatacaatgaagatcagtc	541		
OY 2249	aattctacttaccaatggaaagatcac--aagcagtgtaagatggttcacaggtattactg	2305		
DB 542	ctttctacogtctatgtccaataaaaaatccgaagccacaagaagtgttcgcgaaggtattcctgt	601		
OY 2306	gtacacatgtatgaaagaagtgtccagggagcagcgtgttaccaccaaagaatgacatcattca	2365		
DB 602	gtagaataagagtattataagtgtccaaggagcagcgtctcttaagtatgagatgicagaattgtg	661		
OY 2366	ataaagtacagagactctatagaanaagaatgttgagtgttgcctccaatccgcagacggt	2425		
DB 662	attccacaaaactcgtatggaaaagaattgttgcattcttcttcataaagtacaaacag	721		

QY	2426	agagagcctctctataatggtttgccaacaacggttgatctctatgtatgtaatctctgtacaac	2485
Db	722	aaaaagcatccataataatgtttatgtcaaaagatgatctctgttctgttaattctgtacaagaa	781
QY	2486	aaaaaccacaanaaagaagctcccaacaagaacaaatacaaaatgacatctccgaagacat	2545
Db	782	aaaccataataccaagaagctcccaagccttacaanaacataaagtgcaatttttgaagaatcat	841
QY	2546	ggagatgtatctcgttgatctctgagacctttaaagaaaaccctctatgtacaacacagccac	2605
Db	842	ggagagtgatataagtaattccgtgagatgttttaaaaacacatacccatgtgtgacacacactc	901
QY	2606	caaatcccaaccttctatctgtctgcagaattgagacaagaatgtgtgtttatgtccttgaca	2665
Db	902	ctccacatgtcttctctcatctgtctaaagatcagtaacaagaatctgtgtcttaattgtatga	961
QY	2666	aatccggaagcatalgagcagctgttaacccgctccatccagatgataccaagaagcagacctt	2725
Db	962	agctctgaaagataggggggttaagagaccgcgcttaacatccgaatgatacagcagcaaaactt	1021
QY	2726	tcctctctgcagacagatgtgagctctgggttcctctgggtttggagatgtgacattgtacagctg	2785
Db	1022	tcctctctgcagcagctgttgaaaatgagatcctctgggtggagatgttcaattgtagtaactg	1081
QY	2786	cccatgtacaanaatgaaactataacagataaacaagctgacgtgacaagggacaacctgcaca	2845
Db	1082	cccatctgttataataaagctctcatatccaaataaagaacagtgatgttaagaagaacacactctg	1141
QY	2846	aaagatctacccgcagcagctctcagagggagacgtccatccctgcagcgggtcttcagctcagat	2905
Db	1142	cagagatctaacctacatccctctggaggaacttccatccctgcagctcgtgaaatataatgcat	1201
QY	2906	ttaactgatalgtgtgcacaattttgcctgttctccatgtacacacagcagtgatagggagatgc	2965
Db	1202	ttcaggtgattggagagc-----	1219
QY	2966	gacaagaanaatccaatgtgacctctctgcgtcgtcagcttgatgtagtagaagaanaatc	3025
Db	1220	-----taaatcc	1228
QY	3026	caactgtatgcatctgtaaatgtgctgtcagatgagatgggaagaacaactataagtggt	3085
Db	1229	aactcgtatggtatccgaagctacgtctgcgtcagatcgtgagggagagataaacgcgtcaegtctc	1288
QY	3086	gctttaaagagcttcaaaaacaagaatgtgtgcacatccacaacagatcgtcttggggccctctg	3145
Db	1289	gtattgtatgaaagtgaacaacaagtggggccatgtgtcatcttattcttcttgggaagagctg	1348
QY	3146	cagctcaagaactagagagagctgtcccaaaatgacaagaggtttacagacataatgctcag	3205
Db	1349	ctgatgtgaacgatatagagatatgacaagaataacagagatcatcttattatgtttcag	1408
QY	3206	atcaagttcagaaaanaatgsgctcatgtgacttttggggcccttcatcagaagaatggag	3265
Db	1409	atgaaagctcagaanaaavggccatctgactgtcttcttgggctcttaacaacaggaataatctg	1468
QY	3266	ctgtctctcagcgtctccatccagctgtgagatgaagagatlaaacctccagaacaagcagct	3325
Db	1469	atcctctccagaagatccctctcagctctcgaagaatgaagagatlaaacactgtatgtaatgct	1528
QY	3326	ggatgaatgtgacagatgatctgtgacagaacgctvggaagaacacttgttctctatca	3385
Db	1529	ggatgaacgacatgctcataataatgtatgatacagttgggaagaacacgcttcttctcatca	1588
QY	3386	ccgtgacaatgtgcagctctccccaatactctctctctcgggattcccgatgacagaagaagtgt	3445
Db	1589	catgtgaacagctctgtcctccagatattctctctcgtggtatcccgatggaacaataatgtgaa	1648
QY	3446	gcttgtgagtgtggaacaaaacaacaanaatgtgcttaccatcccaatccagagatctgaag	3505
Db	1649	atttcaacagtgtgagtgaacttccaaaatgtgcttaccatccagtatcccaaggaactgtaaaag	1708
QY	3506	tttgacattgtgaaatacaagctctgc-----aagcaagctccacaanaacctgtgacctgtagct	3559

[illegible]

XX	DT	15-MAY-2001	(first entry)	
XX	DE	Human PRO1124 cDNA.		
XX	KW	Human; PRO protein; mapping; ss.		
XX	OS	Homo sapiens.		
XX	PN	WO200116318-A2.		
XX	PD	08-MAR-2001.		
XX	PF	24-AUG-2000; 2000WO-US233328.		
XX	PR	01-SEP-1999; 99WO-US20111.		
XX	PR	15-SEP-1999; 99WO-US21090.		
XX	PR	07-DEC-1999; 99US-0169495.		
XX	PR	09-DEC-1999; 99US-0170262.		
XX	PR	11-JAN-2000; 2000US-0175481.		
XX	PR	18-FEB-2000; 2000WO-US04344.		
XX	PR	18-FEB-2000; 2000WO-US04342.		
XX	PR	22-FEB-2000; 2000WO-US04414.		
XX	PR	01-MAR-2000; 2000WO-US05601.		
XX	PR	03-MAR-2000; 2000US-0187202.		
XX	PR	25-APR-2000; 2000US-0199397.		
XX	PR	22-MAY-2000; 2000WO-US14042.		
XX	PR	05-JUN-2000; 2000US-0209832.		
XX	PA	(GENTECH ) GENENTECH INC.		
XX	PI	Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;		
XX	PI	Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;		
XX	DR	WPI; 2001-183260/18.		
XX	DR	P-ESDB; AAB87560.		
XX	PT	Eighty four nucleic acids encoding PRO polypeptides, useful in		
XX	PT	molecular biology, including use as hybridization probes, and in		
XX	PT	chromosome and gene mapping.		
XX	PS	Claim 2; Fig 69; 278pp; English.		
XX	CC	The present sequence is the coding sequence for a human PRO polypeptide		
XX	CC	(secreted and transmembrane). The PRO protein, and PRO agonists, PRO		
XX	CC	antagonists or anti-PRO antibodies are useful for preparation of a		
XX	CC	medicament useful in the treatment of a condition which is responsive to		
XX	CC	the PRO protein, agonists, antagonists or anti-PRO antibodies. The PRO		
XX	CC	protein may also be employed as molecular weight markers for protein		
XX	CC	electrophoresis. The PRO coding sequence has applications in molecular		
XX	CC	biology, including use as hybridisation probes, and in chromosome and		
XX	CC	gene mapping.		
XX	SO	Sequence 3265 BP; 1159 A; 596 C; 632 G; 878 T; 0 other;		
QY	Query Match	24.7%; Score 1129.8; DB 22; Length 3265;		
QY	Best Local Similarity	67.1%; Pred. No. 2e-300;		
QY	Matches 1749; Conservative	0; Mismatches 74; Indels. 11; Gaps		
DB	1889	aggaaatgtgtgaccagcagcatctgtatctgttgccttgaagctacaggaagcgatttatt 1948		
DB	182	aggatattgtgtgactacagctcttctacgtcttgaagccacagaaaagatttttt 241		
QY	1949	tcaaaaatgttgcatttttgattcctctgaacatlgaaagacaagaagctgacatgtgaac 2008		
DB	242	tcaaaaatgtatctataatattcctctggaattggaagaaaatctcagtaacaaagc 301		
QY	2009	caaaacttgagacgtcaacaaaatgtgtatgtttcgtttgcctgagctactactccaggta 2068		
DB	302	caaaaatgtgaaacaacataaacatgtctgattgttatgtttgcaccacctcaccaggta 361		
QY	2069	atgacgaacccctacactgagcagatcgtggcgaactgtgagagagaaggttgaagatccacc 2128		

Db 362 gaatgaaacatacaacaagcagttcacagaatgttgagagaagaggaataactact 421  
QY 2129 tcaactcgtatcattcagcaggaanaaagttaagatctgaaatctggaccaggtgagc 2188  
Db 422 tcaacctgacctctactctggaaaaaacaataatgaaatctggaccacagcgaactgt 481  
QY 2189 ttgtcgaatgagtggtcactacagatgagggagbatttgaacagatacaatgataga 2248  
Db 482 ttgtccatgagtggtgcacctccggtgaggagtggttgaatgaatacaatgaatcagc 541  
QY 2249 aattctactatccaatgagaataac---aagcagtaaatgttcaacgaggtacttg 2305  
Db 542 ctctctacgctgcaaaatcaaaaaaaatcgaaacaaagtgctccgcaggtactctg 601  
QY 2306 gtaacaaatgtaaaagatgtlcaagggagcagctgttaccacaaagaatgacatcca 2365  
Db 602 gtgaagaatagattataaagtgtcaaggaagcagctgtcttaagtagacatgcaagattg 661  
QY 2366 ataaagtaacagactctatgaaanaagatgtgattgttcttcccaatccgcagcag 2425  
Db 662 attctacaaacaaactgtatgaaanaagatgttcaatcttccgtataaagtaacaaag 721  
QY 2426 agaaagttctataatgtttgcaacaatgttgatctatagttgaattctgtacagaa 2485  
Db 722 aaaaagatccaatgatttataatgcaaaagtattgtctgttgaattgttgaacgaa 781  
QY 2486 aaaaaccaaacaagaagctcccaacaagaacaaatacaaaatccaatctccgaagc 2545  
Db 782 aaacccaatacaagaagcctcccaagcctatacaaaataaagtgaatttgaagaatcat 841  
QY 2546 gggaaatgtacgtgattctgaggaacttaagaanaaccactcctatgacaacagcac 2605  
Db 842 gggaaatgtatgaatcactcggagattttaaaacataccatcagtgtagacccctc 901  
QY 2606 caatccacactctcaatgtctgacagattgaaanaaatgtgtgttgaatcttgaca 2665  
Db 902 ctccacactgtctctcattctgtgaaagcagtcacaagaattgtgtctatgtcttgata 961  
QY 2666 aatctggaacatgagcagctgtgaacgcctcaatcagactgaatcaagcagccagctt 2725  
Db 962 agctctgaagcagtgagggtgaagcagcctaataatgaatcaagcagcaaaaact 1021  
QY 2726 tctgtctgacagagttgagctggttccgtggttgagatgtgacatttgaacgtgctg 2785  
Db 1022 tctgtctgacagctgttgaataatggaatcctggttgagatgttgaatctgtgaatg 1081  
QY 2786 cccatgtacaanaatgaaactacatacagataaacagtgagtgacaggaacacactgcca 2845  
Db 1082 ccaactattgaaataagcctaataccaataaagaacagtgatgaaagaacacactatg 1141  
QY 2846 aaagattacctgacagcagcttcaagggagcagctcactgtgacagggcttgcatactg 2905  
Db 1142 caagattacatactctctgtgaggaactcctcactgtctgtaataatgatcat 1201  
QY 2906 ttactgtatgtgcaacatttgcgttttccatgacaacagcagttatgaggagctg 2965  
Db 1202 ttcaagtgattgagagc----- 1219  
QY 2966 gacaagaanaatccaattgagcctctctgacctgacagcttagtgatlagagaanaatc 3025  
Db 1220 -----taccatccc 1228  
QY 3026 caactgtatgacatgaaattgtgtctgacagatgagggaagaacacactataatgag 3085  
Db 1229 aactcgatgtatcgacgaagtactgtctgtcagtatgaggggagtaaacacgcgaagttct 1288  
QY 3086 gctttaaagagtgcaacaagaatgtgtgcatactcaacacagtcgcttggggcctctg 3145  
Db 1289 gttatgtatgagtgaaacaagaatgagggtgagcattgtatattgttgggagagcgtg 1348  
QY 3146 cagcttaagaacatgaagggtgttccaanaatgacagaagtttaacgaacataagcttcag 3205

Db 1349 ctgataagcagtaataatagatgagcaagatacaacaggaaggaatcatttattgttccag 1408  
QY 3206 atcaagtctcaagaacaatgagcctcattgaltgttlttgggcccttcatcagaanaatggag 3265  
Db 1409 atgaagctcgaacaatgagcctcattgtatgttlttggtctcttaatacagaanaatcag 1468  
QY 3266 ctgtctctcagcctccatcccaatcagcttgagagtaaggatlaaacctccgaacagcag 3325  
Db 1469 atctctccagaagatccctcagctcagtaagaaatgaagatlaaacatgataatgagcct 1528  
QY 3326 ggaatgaatgacagatgtgtgacagacagcttggaagaaggaacattgttcttata 3385  
Db 1529 ggaatgaagcactgttcaataatgtatagatcagtggaagaagacagttcttcttata 1588  
QY 3386 cctggaacatgacagcctcccaaatcctctctcggagatcccaatggaacaggaagcag 3445  
Db 1589 catggaacagctgcctcccaatctctctcggatcccaatggaacaggaatgaaatgaa 1648  
QY 3446 gcttgtatgtgacaanaaacaatgagcttactcccaatcccaagcattgtcag 3505  
Db 1649 atttcaagtgatgtaacttccaanaatgagcctatctcagttatccaaggaactgcaag 1708  
QY 3506 ttgacacttggaataacagctg-----aagcagctccacaacacttgacctgag 3559  
Db 1709 tgggcacttgaggatacaatcttcaagccaagcgaacccagaataactatatacag 1768  
QY 3560 tcaactcctgtgcttcaatgctacccctgctcccaatcaatgagacttccaacagca 3619  
Db 1769 taactctctgagagcaaatctctctgtctcccaatcaacaggaatgttaaatgaaat 1828  
QY 3620 aggaacccagcaaatccccaagcctctgtgattatgtcaaatctgcgaagagcct 3679  
Db 1829 aggaacgtaaacagttccccaagccaatgattgttaccagaabaatttcaacaagatag 1888  
QY 3680 cccaatctcgaaggcagctgtacagccctgatttgaatcagatgaaatgaaacagtt 3739  
Db 1889 taactgttcttgagccaatgtgacgtcttcaatgaaatcaacagatgacatacagaag 1948  
QY 3740 ccttgaactactgataatgagcaggtgtgtatgtctactaagataagcaggtgtact 3799  
Db 1949 ttgtgaactttgtataatgtgagcagcgtgattcttcaagaatgtgtgagctact 2008  
QY 3800 caaggtatttcaactatgacagaatgtgataacagttgtaaaagtgcgggtctg 3859  
Db 2009 ccaaggtatttcaagatatagaaatgtgcagatatagcttaaaagtgcgggtcag 2068  
QY 3860 gaaggtatacgaagcagcagagatgatacccaagcaggtgaggaacttataac 3919  
Db 2069 gaggagcaaacactgtccaggtcaaaatctgcgtcctcactgaatgagccgttataac 2128  
QY 3920 ctggtgtgtatggaatgataaatacaatggaatcccaacgaacctgaaataaag 3979  
Db 2129 caggtgtgtatggaatgataaatacaatggaatcccaacgaacctgaaataaag 2185  
QY 3980 atgatttcaacaacagcaggtgtgttgaagcaacatccctcggagagctatgtgtg 4039  
Db 2186 aggatctcaagccacactgtgagatltcagccgaacacagatccggaggtgtgattgtg 2245  
QY 4040 ctctgtatgtcccaaatgtcccatatcagctcactcccaactgtgccaataccagcc 4099  
Db 2246 tatcaagaatcccaagccttccctgtccgtgacccaatacccaacgaatccaatacagacc 2305  
QY 4100 tgaagcggaaatctcagggggcaggtcataltaactgtgacttgacagctcctggagatg 4159  
Db 2306 ttgattgccaagttcatgag---ataagattatcttcaatgagcagcaagagata 2362  
QY 4160 atatagaacatggaacagctcacaagtatatactcgaataagtaagtaagttatctgac 4219  
Db 2363 atttgaatgttgaaagaattcaaggtatatacataagaataagtaagtaagttatctgac 2422  
QY 4220 tgaaggaacagttcaatgatacttcaagttgaataactactgtctcattcccaagaagag 4279  
Db 2423 taagagcagtttgaatgtgtcttcaagtaataactactgtactgttcaaccaagaagag 2482





Db	782	aaaccataatcaagaagcctcccaagcctacaacaacctaagcgaattttgaagtaac	841
Qy	2546	gggaagtgatccgtgtaattctgaggaactttaagaaaacccctcatagacaacagccac	2605
Db	842	ggagagtgattgaactctgaggaatttttaaaacacatacccatggtgacaacccctc	901
Qy	2606	caaatcccaacttctaatctgtcagattggaataaagtgtgtttagtccttgcga	2665
Db	902	ctccacactgtcttctcatctgtcgaagatctcagtaaagaattgtgtctatgtctctga	961
Qy	2666	aatcctggaagcatgctcagctgtaaacgcctccatctgactatcaagaagccagctt	2725
Db	962	agtcctggaagcagatggtgggtgaaagaccgcctaaatctgattgatacaagcaaacact	1021
Qy	2726	tcctgtctgacagacagcttgagctctggtgtccctgggtctgggaatgtgtacatttgacagtgct	2785
Db	1022	tcctgtctgacagactgtgtgaataatgatactccgtgtgggtgtgttccatttgatgactgt	1081
Qy	2786	cccatgtctcaaatgaaactcatacagataaacaagctgtcagctgacagggacacactgcga	2845
Db	1082	cccatctgttaataatgaactcatccaaataaagaacagctgtatgaagaacacacactcatg	1141
Qy	2846	aaaagttacccctgacagcagctctcagggggagcgtccactctgcagcggtctgtatggat	2905
Db	1142	cagagcttaccacacatactccctcgtgggaagacttccactccctgcgtctggataataatgcat	1201
Qy	2906	ttactgatatgtgcaacaacatttgcctgtttcttcacatgacacagagttatgtggatgctc	2965
Db	1202	ttcaggtgtatgtggagagc-----	1219
Qy	2966	gacaaagaaatccaaattggtgcctctctgtcctgtcagcttgatgtatggagaataatc	3025
Db	1220	-----taacttccc	1228
Qy	3026	caactgattggaactgtaaatctgtctgtctgaagatggtgggaagacacactaatgaatggt	3085
Db	1229	aactcgtatggtctccgaaagatctcgtctgtcagctggaatgggaagataacactgcaagcttct	1288
Qy	3086	gctttaagaggtlcaaaaacaagaatgtgtgccatcatccacacagctgccttcttgggcccctgt	3145
Db	1289	gtattgtatgaagtgaacaacaagctggggccactgttcatcttattgtcttgggaagagctg	1348
Qy	3146	cagctcaagaagactgagagagctgtccaaaaatgacagaggtttacagacatatgctcag	3205
Db	1349	ctgtatgaagcagtaataagatagagacagaataacagaggaagatcaattttatgtttcag	1408
Qy	3206	atcgaagtccaanaaagtgccctatgtatgcttcttgggccccttcatcagaagaatggag	3265
Db	1409	atgaagctcagaaaaaatggtcccatatgtgacttcttgggccccttcatcagaagaatactg	1468
Qy	3266	ctgtctctcagcgtctccatccacagctctgagatgaagatataccctccagacagacagct	3325
Db	1469	atctctccagaagtctcccttcagctctgaaagatgaagtgatatacactgtatgtatggct	1528
Qy	3326	ggatgaatgtagcagtgatctgtgtgacagcagctgtggaaagacacttgtttcttatca	3385
Db	1529	ggaatgaacagacatgctataatgtatgatacagctgtggaaagacacgcttcttctcatca	1588
Qy	3386	ccctggaacaatgtagcgtccctcccaaatctctctcgtggatcccgatgtagacagaagcaagtg	3445
Db	1589	catggaanaagctctgtcctccagatattctctctgtggtatcccgatggaacaataatgyna	1648
Qy	3446	gcttctgtatgtggaacaaaacacaaaatggtccctaccctcaaatcccaagatgtctaag	3505
Db	1649	attccacagtgtagtgcgaacttccaaaaatggtcctatctcagttatccaggaactgcaaa	1708
Qy	3506	ctggcacttgggaataacagctctgc-----aagcaagctccaanaaccttgaccttgatg	3559
Db	1709	ctggcacttgggcataacatcttccaaagccgaagccgaanaacatttaactttaag	1768
Qy	3560	tcaagctccggtgtgctccaatgtctacccctgtcccaattacagtgacttccaanaagaca	3619
Db	1769	taactctctgagcagcgaatacttctctgtgtccctccaatccagctgtaatgtctaaatga	1828

QY	3620	aggacacccagcaaatlccccagcccttcgttagttatgcaaatlcccgaaagagcc	3679
Db	1829	aggagatgaataacagttlccccagcccaatgaatttcttacgcagaataatctcaaatgagatg	1888
QY	3680	cccacatcttcaggccagtgatccacgcccctgtatgaatcagtgatggaanaacgta	3739
Db	1889	taccgttcttgaggcccaatgtagtcgcttctcaatgaaaccaagaatgagatacagaag	1948
QY	3740	cccttgaaactactggaataatgagcaggtgctgtagtgcctactaagaatgacggtgctact	3799
Db	1949	tttttgaaacttcttgataatagtgtagcgctgctatctcttccaagaatgatgagttctact	2008
QY	3800	caagatatttcaacaactatgacacgaatgtagatagaagtgtaanaagtgggggcttcg	3859
Db	2009	ccagatatttlltcaacgcatatcacgaaatgtagcagatatgcttaanaagtccgggcttcg	2068
QY	3860	gagagatgaacgcagccagacgtagagatgataccccagagagtgagacactgtataac	3919
Db	2069	gagagagcaaacacgcgcaggtctaaataatgaagctctccactgataagcgcgtataac	2128
QY	3920	ctgctgtagattggaataatgataatatacgaatgcgaatcccaagacctgaatataga	3979
Db	2129	caggtctggttagtgtaacgggggaatattgaacaaacccgcgaagacctgaatatgat--g	2185
QY	3980	atgagatgcaacaacaagaacagtgctgttctcagcaaacatccctcgggagggctcattgtg	4039
Db	2166	aggaataccagacacacacttggagattctcaagccgaaacagcatccgagagtggtcattgtg	2245
QY	4040	ctctctgagtcaccaaatgctcccaatacctgatactcttcccacgtgcgaataatccagac	4099
Db	2246	latccaaagttcccagagcttcccttcttcgtctatcccaatacccaacgaagttcaatccagacc	2305
QY	4100	tgaaggcgaatctcacgcgggggcagctccatgaatcttgacttgacttggaagctcctctggagtg	4159
Db	2306	tgtgagccacagcttcatgagag---ataagaatttcttcatgagcagcagccagagagata	2362
QY	4160	attatgaccatggaacagctccacaagtatatacttcgaataagtaagtaattcttgatc	4219
Db	2363	attttgagtttggaanaaggttccaagttatatactcaagaataagtgcaagttacttctgac	2422
QY	4220	tcagaggaacagttaacgaatctctccacagtgatactactctgctctcaatcccaagaagag	4279
Db	2422	taagagacagattttagatagtcctcccaagtaatactactgtagtctgcccaccaagaagag	2482
QY	4280	ccaactctgaggaagctctttgttttaaacccagaanaacatacttcttganaatgtagcag	4339
Db	2483	ccaactccaaggaagagtttgcatltaaacccagaanaatactccagaagaaatgaaoccc	2542
QY	4340	atcttcttcattgcattcagcgtctgtataagtgtagatctggaatccagaataatccaca	4399
Db	2543	acatatattatgccaataaagaatagatahaaagcaatttgcatacaaaagatltccaca	2602
QY	4400	ttgcaggaatactctgtttatctccccaagaatccgcgcaaga---cacctagtcctg	4456
Db	2603	ttgcacaagaatacttgtttatccctcccaagaataatccgtatgacattgacttactaacta	2662
QY	4457	atgaaacgctgcgtccttgcttactaata	4483
Db	2663	ctccctacactcactccactccactgata	2689

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 12:59:18 ; Search time 8456.85 seconds

(without alignments)  
8912.966 Million cell updates/sec

Title: US-09-867-034-3

Perfect score: 4569  
Sequence: 1 atcgataaattcctcaacac.....tgcagctgcataagcctag 4569

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_om:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rod:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2492.2	54.5	3311	6	AX193489	AX193489 Sequence
2	2492.2	54.5	3311	9	AF039400	AF039400 Homo sapi
3	2473.6	54.1	2826	9	AF127036	AF127036 Homo sapi
4	1847.2	40.4	2022	9	AK024970	AK024970 Homo sapi
5	1637.8	35.8	3079	4	AF095584	AF095584 Sus scrofa
6	1517.6	33.2	2937	10	AB017156	AB017156 Mus muscu
7	1136	24.9	3204	9	AF127035	AF127035 Homo sapi
8	1129.8	24.7	3221	9	AK000072	AK000072 Homo sapi
9	1129.8	24.7	3265	6	AX092338	AX092338 Sequence
10	1058.2	23.2	175591	2	AC068071	AC068071 Homo sapi
11	1051.8	23.0	140718	2	AL356270	AL356270 Homo sapi
12	1049.4	23.0	118426	2	HS431C21	282210 Human DNA s
13	1033.8	22.6	196805	2	AC010275	AC010275 Homo sapi
14	1024.8	22.4	168185	2	AC079805	AC079805 Homo sapi
15	1005.4	22.0	149428	2	AC010264	AC010264 Homo sapi
16	1005.4	22.0	186882	2	AL136307	AL136307 Human DNA
17	1005.4	22.0	187045	2	AC073253	AC073253 Homo sapi
18	1004	22.0	151200	2	AC015658	AC015658 Homo sapi
19	1004	22.0	154712	2	AC010958	AC010958 Homo sapi
20	1003.8	22.0	99074	2	HS560B9	298751 Human DNA s
21	1003.8	22.0	151551	2	AC021291	AC021291 Homo sapi
22	1003.8	22.0	152408	2	HS191J18	AL024507 Human DNA
23	1003.8	22.0	157670	2	AC074331	AC074331 Homo sapi
24	1003.8	22.0	170425	2	AC024490	AC024490 Homo sapi
25	1003.8	22.0	176087	2	AC024625	AC024625 Homo sapi
26	1003.8	22.0	186780	2	AC005740	AC005740 Homo sapi
27	1003.8	22.0	187766	2	AC016042	AC016042 Homo sapi
28	1003.8	22.0	197927	2	CNS05TE0	AL158293 Human chr
29	1002.4	21.9	126039	9	AC012665	AC012665 Homo sapi
30	1002.4	21.9	168937	2	AC018747	AC018747 Homo sapi
31	1002.2	21.9	110000	2	AL360016_1	Continuation (2 of
32	1002.2	21.9	130244	9	AL590705	AL590705 Human DNA
33	1002.2	21.9	150011	2	AC068232	AC068232 Homo sapi
34	1002.2	21.9	157832	2	AC011653	AC011653 Homo sapi
35	1002.2	21.9	182938	2	AL512304	AL512304 Homo sapi
36	1002.2	21.9	196934	9	AL355530	AL355530 Human DNA
37	1002.2	21.9	197274	2	AL357373	AL357373 Homo sapi
38	1002.2	21.9	211544	9	AC025165	AC025165 Homo sapi
39	1000.6	21.9	70113	9	AC002478	AC002478 Human BAC
40	1000.6	21.9	122272	9	AL359845	AL359845 Human DNA
41	1000.6	21.9	139737	9	HS228H13	AL031985 Human DNA
42	1000.6	21.9	141320	2	AC074322	AC074322 Homo sapi
43	1000.6	21.9	148098	2	AL359315	AL359315 Homo sapi
44	1000.6	21.9	150983	2	AL138802	AL138802 Homo sapi
45	1000.6	21.9	159160	2	AC018624	AC018624 Homo sapi

## ALIGNMENTS

RESULT 1  
AX193489 3311 bp DNA PAT 15-AUG-2001  
LOCUS AX193489 Sequence 1056 from Patent WO0149716.  
DEFINITION AX193489  
ACCESSION AX193489  
VERSION AX193489.1 GI:15211440  
KEYWORDS  
SOURCE  
ORGANISM human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,  
Stolk,J.A., King,G.E., Wang,T., and Jiang,Y.  
Compounds for immunotherapy and diagnosis of colon cancer and  
methods for their use  
Patent: WO 0149716-A 1056 12-JUL-2001;  
JOURNAL CORIXA CORPORATION (US)  
TITLE  
AUTHORS  
FEATURES  
SOURCE  
1. 3311  
Location/Qualifiers

BASE COUNT	1028 a	692 c	742 g	849 t
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606"			

Query Match	54.58;	Score 2492.2;	DB 6;	Length 3311;
Best Local Similarity	91.3%;			
Matches 2772; Conservative	0;	Mismatches 8;	Indels 255;	Gaps 2;

QY	1535	tgcccaagagagaaatcaacsaggagatgacagaaatgggccaattaaagtlctcgt	1594
Db	317	TTCCAAAGAGAGAGATCACAGGGGAGATGTACACCAATGGGGCCATTAGAGTTCTGAGT	376
QY	1595	tcatttgattcttcacctctctagaaggggcccttgatgaatttcattcattcgtcgaca	1654
Db	377	TCATCTTGATTTCTTCACCTCTTGAGAAAGGGCCCTTGAGTAATTCACATTCACAGCTGAACA	436
QY	1655	acaaatgctatgaaagcatttgcgttgcaatcgaccaccaatgtgccagaagaatgaac	1714
Db	437	ACAAATGGGTATGAAGGAGATTGTGGTTGGCAATGCAGCCCAATGTGCCAAGAAGATGAACAC	496
QY	1715	tcattcaacaataaaggggagttacacgttcacacagaatgaaggaggagatcagaga	1774
Db	497	TCATTTCAACCAAAATA-----	510
QY	1775	aacctctcttcccccgtaaatataatacacacacacacacagcacagctcgtgtgc	1834
Db	511	-----	510
QY	1835	acacacacacgcccattgacacacgcagacatcacgcacacagcaacgttcagaagaca	1894
Db	511	-----AAGGACA	517
QY	1895	tgtgtgacccagggatctctgtatctgttgaaagctacagaaagcgattatttcca	1954
Db	518	TGGTGACCCACAGGCATCTCTATCTGTTTGAAACCTCACAGAAAGGAGATTTTATTTTCAAA	577
QY	1955	atgttgcatttgattcttcgaaacatggaagcaaggctgacctgtggaagccaac	2014
Db	578	ATGTTGCCATTTTGATTCCTGGAACATGGAAGACAAAGGCTGACTATGTGAGACCAAAAC	637
QY	2015	ttagagacctacaanaatgctgatagtctcgtgtgtgctgagctactcctccagtaata	2074
Db	638	TTGAGACCTCAAAAAATGCTGATGTTCTGGTTGCTGAGTCTACTCTCCAGGTATGATG	697
QY	2075	aacctactctgagcagatgaggaaactgtgaggagaaggatgaagatccaactcactc	2134
Db	698	AACCTTACACTGACACAAATGGGCACTGTGGAGAGAGGGTGAAAGATCCACCTCACCTC	757
QY	2135	ctgattcatttgcaaggaaaaaagttagctgaatcatgagcacaaagttgagcatttgc	2194
Db	758	CTGATTTTCATTGCAGGAAAAAATGTAGCTGAATATGGACCAAGGTAAAGCATTTGTCC	817
QY	2195	atgagtgagctcatctacgaatggggagatattgacagatataatgatagaatctc	2254
Db	818	ATGAGTGGGCTCACTACATCTGATGGGGAGTATTTGACAGATACAATGATGAGAAATCT	877
QY	2255	actatccaatggaagaataatacaagcagtaagatgttcagcaggtattactgttacaat	2314
Db	878	ACTTATCCAAATGGAAGAAATACAGCAGTAAGATGTTCAGCAGAGTATTTCTGGTACAAATG	937
QY	2315	tagtaagaagatgtcaaggaggagctgtgtacaccaaaagatgagacattcaataa	2374
Db	938	TAGTAAAGAAAGTGTCAAGGAGAGAGCTGTTTACACCAAAAGATGCAATTCAATATAGTTA	997
QY	2375	caggactctatgaaaaagagatgtgagttgttctccaatcccccgacagcgagaaagc	2434
Db	998	CAGAGCTCTATGAAAAAGAGTGTGAGTGTGTTCTCCAAATCCCGCCAGAGGAAAGGCTT	1057
QY	2435	ctataatggtttgacacaacatggttgaattctcatggttgaattctgtacagacaacaca	2494
Db	1058	CTATTAATGTTTGCACAACAAGTTGATTCTATGTTGTAATTCGTGTACAAACAAAAACCA	1117

QY	2495	acaaagaagctccaaacaagcaaaatcaaaatgcaactctcgaagcaca	tgaggagtga	2534
Db	1118	ACAAAAGAGCTCCAAACAGCAAAATC	AAAAATGCAATCTCCGAACCATGTGGAAATGA	1177
QY	2555	tcctgattctggaagactttaagaanaacactctctatgaanaaagacccaatctcca		2614
Db	1178	TCCGTGATTCTGAGGACTTTTAAAGAAAACCACTCTTATGACACACAGCCACCAATCCCA		1237
QY	2615	cccttcactgctgcgaattggaacaaagaattgtgtgttagtctcttgacaaatctgaa		2674
Db	1238	CCTTCTCATTTCTCTCAATGGATGGACAAGAATTTGTGTTTAACTTCCTTGCATCAATCTGGAA		1297
QY	2675	gcattgagcagctggttaaacccgctcaatctgactgatacaagagccagcttcttcgtctgc		2734
Db	1298	GCATGCGCACAGTGAACCGCTCAATCGATCGATGATCAACAGAGCCAGCTTTTCGTCGTCG		1357
QY	2735	agacagcttgaagcttgaggctccgtggttggtgaatgaatctgacattgacgtgctgcatactgac		2794
Db	1358	AGACAGTTGAGCTGGGGTCCGTGGGTTGGATGGTGAATTTGACAGTGTCTGCCATGTAC		1417
QY	2795	aaatagaaactcatcagataaacaagctgagctgaacagggaacacactgcgcaaaagaattac		2855
Db	1418	AAAGTGAACCTCATTCACATTAACAGTGCAGTACAGGAGACACACTCGCAAAATATAC		1477
QY	2855	ctgcagcagctctcaaggagggaagctccaatgcagagcgagcttcgatctgacattactgata		2914
Db	1478	CTGCAAGCAGCTTCAGAGAGGAGACGTCCATCTGCAGCGGGCTTGATCAGCATTTACT		1533
QY	2915	tgtggaacaacttgctcgtgttctccatgaacacagcagttatggtgagtgctgcagaagaa		2974
Db	1534	-----		1533
QY	2975	atccaattgggcctctctgacctgagcttagtgaattggagaataatccaactgata		3034
Db	1534	-----GTGATTAGAGAAATATCCAACTGTAG		1561
QY	3035	gactgaaattgtgctgctgcagatgagtgaggaaagacaacactaaagtgtgtgtctttaaag		3094
Db	1562	GATCTGAATTTGTCTGCTGCTGACGATGGGAAACAACTAATAGTGGTCTCTTAAAG		1621
QY	3095	aggttcaaaaagaagtggtgcataatccaacagcttgagctttggggccctctgtagacttaag		3154
Db	1622	AGGTCAAAACAAGTGGTGCATATCCACACACTCCCTTTGGGCCCCCTCTGCAAGCTCCAG		1681
QY	3155	aactagaggagctgtcccaaaatgacaggaaggtttacagacatatgcttcaagatacgaatgc		3214
Db	1682	AACATAGAGGAGCTGCCAAAATGACAGAGGTTTACAGACATATGCTTCAGATCAAGTTC		1741
QY	3215	agaaacaatgacctcatatgtatgtctttggggcccttcatacaggaataatgtagctgtctctc		3274
Db	1742	AGAACATATGGCTTCATTTGATGTCTTTTGGGGCCCCCTTTCATCAGGAAATGGAGCTCTCTCTC		1801
QY	3275	agagctccatccagctcttgagagatlaaaggaattaaacctccagaagacagcagtgatgaaatg		3334
Db	1802	AGGCGTCCATCCAGCTTGAAGATTAAGGATTAACCTCCAGAAACCCAGTGGATGATG		1861
QY	3335	gcacagctgatactgagacagcagcttggaagaagacacttgttctcatcaactctggaaca		3394
Db	1862	GCACAGATGATCGTGGACAGCACCGTGGGAAAGACACTTGTTCCTTATCACTGGACAA		1921
QY	3395	tgcagctccccaataatctctctctgtggaatcccaagtggaagaagaagaagtgagctttgag		3454
Db	1922	CGCAGCTCCCAAAATCTTCTCTGGGATCCCAAGTGACAGAGCAAGGATGGCTTTGAG		1961
QY	3455	tgagcaaaaaacccaataatgctctactccaatcccaagcatgtgtlaaggttgtagctt		3514
Db	1982	TGACACAAAACACCAAAATGTGGCTTACTCTCAATCCAGACATTCCTAAGGTGGCACTT		2041
QY	3515	ggaataacagctctgcaagcagctcaaaaacttgaaccttgactgttcaagctccgtgctgt		3574
Db	2042	GGAAATTCAGCTCTGCACAGCAAGCTCCAAACCTTGAACCTCTGACTCTCAAGTCCGCTGGCT		2101

```

Oy 3575 ccaatgctacccctgcctccaattacagtgacttccaaaacgaagaacacgaacaaat 3634
    |||
Db 2102 ccaatgctacccctgcctccaattacagtgacttccaaaacgaagaacacgaacaaat 2161
Oy 3635 tcccaagccctcgtgtaattatgtaaatatcgcgaaggagccctcccaattccagg 3694
    |||
Db 2162 tcccaagccctcgtgtaattatgtaaatatcgcgaaggagccctcccaattccagg 2221
Oy 3695 ccaagtctacagcccttgattgtaacagtgaaatgtaaaaacagttacattggaactctg 3754
    |||
Db 2222 ccaagtctacagcccttgattgtaacagtgaaatgtaaaaacagttacattggaactctg 2281
Oy 3755 ataatggaagcagtgctgtaactgtaagatgacggtgtctactcaagtaattcaaca 3814
    |||
Db 2282 ataatggaagcagtgctgtaactgtaagatgacggtgtctactcaagtaattcaaca 2341
Oy 3815 ctcttgacaagaatgctgtaactgtaagatgacggtgtctactcaagtaattcaaca 3874
    |||
Db 2342 ctcttgacaagaatgctgtaactgtaagatgacggtgtctactcaagtaattcaaca 2401
Oy 3875 ccaagcaggaagtgatgtaacccagcagatggaacactgatactgactgactgactgag 3934
    |||
Db 2402 ccaagcaggaagtgatgtaacccagcagatggaacactgatactgactgactgactgag 2461
Oy 3935 atgatgaaatcaaatggaatcccaacgaacactgaaataataaggaatgattcaacaca 3994
    |||
Db 2462 atgatgaaatcaaatggaatcccaacgaacactgaaataataaggaatgattcaacaca 2521
Oy 3995 agcaagtgcttgtaagagaacatctcgtggaaggtctatcttggtctctgactgactc 4054
    |||
Db 2522 agcaagtgcttgtaagagaacatctcgtggaaggtctatcttggtctctgactgactc 2581
Oy 4055 atgctcccaatccctatctcttcccaactggaacatcccaacgaacactggaagcgaatct 4114
    |||
Db 2582 atgctcccaatccctatctcttcccaactggaacatcccaacgaacactggaagcgaatct 2641
Oy 4115 acgggggagctcctcaatcactgtaactggaacagctcctcgggagtgatgatacgaatg 4174
    |||
Db 2642 acgggggagctcctcaatcactgtaactggaacagctcctcgggagtgatgatacgaatg 2701
Oy 4175 cagctcacaagaatatactcttgaaataaagtaaatctcttgatctcgaagagaatgctca 4234
    |||
Db 2702 cagctcacaagaatatactcttgaaataaagtaaatctcttgatctcgaagagaatgctca 2761
Oy 4235 atgaatcctcgaatgataactgctctcatcccaaaaggaacgaacactggaaggaag 4294
    |||
Db 2762 atgaatcctcgaatgataactgctctcatcccaaaaggaacgaacactggaaggaag 2821
Oy 4295 tcttttgcttaaacacgaataaacttacttctgaaatggaacagatcttctcatctgcta 4354
    |||
Db 2822 tcttttgcttaaacacgaataaacttacttctgaaatggaacagatcttctcatctgcta 2881
Oy 4355 ttcaggcgttgtaagtgatcgaatcgaataaacttcaaacatggaacagatcttct 4414
    |||
Db 2882 ttcaggcgttgtaagtgatcgaatcgaataaacttcaaacatggaacagatcttct 2941
Oy 4415 tgttatctccctcgaagctccgcaagagaacactgactcgtatgaataaagctctgctct 4474
    |||
Db 2942 tgttatctccctcgaagctccgcaagagaacactgactcgtatgaataaagctctgctct 3001
Oy 4475 gtcccaatataatcaatacagaacacacttctcgtgcatcacaatttaaaatlatgltga 4534
    |||
Db 3002 gtcccaatataatcaatacagaacacacttctcgtgcatcacaatttaaaatlatgltga 3061
Oy 4535 agtgatagagaagcagctgactgctgtaataagcttag 4569
    |||
Db 3062 agtgatagagaagcagctgactgctgtaataagcttag 3096

```

```

RESULT 2
LOCUS AF039400 3311 bp mRNA PRI 14-DEC-1998
DEFINITION Homo sapiens calcium-dependent chloride channel-1 (hCLCA1) mRNA,

```

```

ACCESSION complete cds.
AF039400
AF039400.1 GI:4009457
VERSION
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 311)
AUTHORS Gruber,A.D., Elble,R.C., Ji,H.L., Schreuer,K.D., Fuller,C.M. and
Pauli,B.U.
TITLE Genomic cloning, molecular characterization, and functional
analysis of human CLCA1, the first human member of the family of
Ca2+-activated Cl- channel proteins
JOURNAL Genomics 54 (2), 200-214 (1998)
MEDLINE 99047526
REFERENCE 2 (bases 1 to 311)
AUTHORS Gruber,A.D., Elble,R. and Pauli,B.U.
TITLE Direct Submission
JOURNAL Submitted (19-DEC-1997) Pathology, College of Veterinary Medicine,
Cornell University, Ithaca, NY 14853, USA
FEATURES
Source
1..3311
Location/Qualifiers
1..3311
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="1p22-p31"
/tissue_type="small intestine"
1..3311
/gene="hCLCA1"
/translation="MGPEKRSVFLLILHLLEGALNSLIOLNNNGYEGIVVAIDPNP
EDFLIOIKDMVMOASLYLEAFATGKRRYEFNVAILIPETMKTRADYVRRLPYKNA
DIVIAESTPRENDPEPYROMNGCEBERHILITDFIAGKLAIEYPOGKAFVEMAH
LRMGVFEDYNDEKFLYSNGRIQAVRCSAGITGVNVAKCGGSCYTRCTFNVTGL
YKGECEFLVDSRQTEKASIMPAQVDSIVERCTQDNHKKKPNQONCNRSTWEYI
RDSBDFKRTTTPMPYOPNPPEFLSLOIGRIYCLVLDKSGSATGNRLRLNOAGLEL
LOVELGSWMGMVTFDSAAHYOSELIQINSQSDBDTLAKRLPAAAGGTSISGLRSL
FTVIRKKYPTGSEIVLITDGEENTISGCEFWEROSQALITHTVALGPSAQDELELSK
MTSGLOPTYASQVONNGLIDFGLSSNGAVSQRSQIILEKGLTLDNSQNMKTIVY
DSIVGKDTLELITWTTPQILLMDPSQKQGFVVDNNTMAVLIQIPGAKVETWIK
SLQASQTLTLTVTSRASNAFLPIITVYSKINKDPSKPPSLVYVANRQOASPIRLA
SVTALIESVNGKVTYLELDNGACADATKDDGVSRFTYDTNGRYSVKRALGVN
AARRVAPIQOOSGALYIPGWLINDEIOMNPPEEINKDOVORHKOYCFSTSGSEFAS
DVNPAPIPDLEPPQIDIDKAEIHGSLINTWTAPAGDDVGHTAHYIIRISYIID
LRQFNSIQVNTYALIPKREANSEVFLFKRENTFPNGDTLFAIQAVDVKSEI
SNIRVSLFIPQTPETPSPDETSAQPCNPHINHSITPIGHILKIMKMGIELDLSTA
"
BASE COUNT 1028 a 692 c 742 g 849 t
ORIGIN
Query Match 54.5%, Score 2492.2; DB 9; Length 3311;
Best Local Similarity 91.3%; Pred. No. 0;
Matches 2772; Conservative 0; Mismatches 8; Indels 255; Gaps 2;
Oy 1555 tggccaagaaggaatacacaagggagaatgatacagaagaatgggacatttaagagttctgtgt 1594
    |||
Db 317 TTCAAAAGAGAGGAAATCACAGGAGAGATGTACAGCAATGGGCCATTGTAAAGATTCTGTGT 376
Oy 1595 tcaatctatcttcaactctagaaggagccctgagtaattcaactcaatcaagctgacaca 1654
    |||
Db 377 TCATCTTATTTCTTACCTTTAGAGAGGCCCTCGAGTAATTTCACTCTTTAGCTGGAACA 436
Oy 1655 acaatgctatgaaagcattgctgtgcaatcgaccccaatgctgacagaagaatgaacac 1714
    |||
Db 437 ACAATGCTATGAAGGCACTGTGCTTGCATTCGACCCCAATGCTGCGACAGAAGATGAACAC 496

```

OY	1715	tcattcaaaataaanaagggagttacaagtcloacaagaatigaaggaagggagtcloagaaga	1774
Db	497	 TCATTCAACAATAA-----	510
OY	1775	aactctctctcccccgtlcaaatatacaacacacacacacagcagacaagctcgtgtgc	1834
Db	511	-----	510
OY	1835	acaacacacagcccaatgcacacacagacacatacacgcacacagcagctcagaagaca	1894
Db	511	-----       -----AAGGACA	517
OY	1895	tcgtgaccaccaagcactctctatctcgtctgaagctccacaggaagagatttatctcaaa	1954
Db	518	TGGTGACCCAGGCAATCTCTGTATCTGTTTGAGACTACAGAAAGCGAATTTTATTTCCAAA	577
OY	1955	atgttgccattcttataccctcgaacaatlygaagaacaagagcttactatlygaaccaaac	2014
Db	578	ATGTTGGCACTTTTGTATTCCTGAACAACATGGAAGACAAAGGCTCATATGTGAGACCAAAAC	637
OY	2015	ttgagaacctacaanaatgtctgtagtcttggttgcgtgagtcctactctccagttaatgtg	2074
Db	638	TTTGAGACCTACAAAAAGCTGAGTGTCTGGTTGCTGAGCTCACTCCTCAGATATGATG	697
OY	2075	aaccttaacatgaaacaaatgggcaactgtggaagaagaagtgaaagagttcaaccctcactc	2134
Db	658	AACCTTACACTGACACAAATGGCACTGTGAGAGAGAGGGTAAAGATCCACCTCACTC	757
OY	2135	ctgaattcattatgcaggaanaaaglttagctgaaatatabgacacaaagttagggcatctgccc	2194
Db	758	CTGATTTTCATTGCGAGAAAAAGTTAGCTGAATATGTGACACACAGSTAAAGCATTTTGCC	817
OY	2195	atgagttggctcatctcctcgaatggggagatattgagagtaacaataatgataagaaattct	2254
Db	818	ATGAGTGGGCTCATCTCGATGGGGAGTATTTGACAGATCAATATGATGAGAAATTCCT	877
OY	2255	acttatccaatggaagaatacaagcagatlaagatgtctcagcaggtatctagttaacaatg	2314
Db	878	ACTTATTCCAATGGAAGAAATACAAAGCATTAAGATGTTTACAGAGTATTTACTGGTACAAATG	937
OY	2315	taataaagaatgttcaaggagagcagctgttatacaccaaagaatgacatltcaataaagtaa	2374
Db	938	TAGTAAAGAAATGTGCAGGAGGAGCAGCTGTATACCACAAAGATGACATTAATAAGTTA	997
OY	2375	caggactctatgaaaagagttgtgagtttggttctccaatcccgccaagcaggaagagctt	2434
Db	998	CAGGACTCTATGAAAAGAGATGTGAGTTGTTCTCCAATCCCGCAGACGGGAGAGGCTT	1057
OY	2435	ctataatgttttgacaacaatgttgattctctatgttaattctctgtacaagaacaacacca	2494
Db	1058	CTATATATTTTGCACAAACATGTGTTCTATTACTTAATTTCTGTACAGACAAACACACA	1117
OY	2495	acaaagaagctccaaacaagcaaaatcnaaanaatgcaactctccgaagcaatatggaagtga	2554
Db	1118	ACAAAGAAGCTCCAAACAAACCAAAATTCAAAATGTCAATCTCCGAAGCACTATGGAGATGA	1177
OY	2555	tcogtgaattctggaagactttaagaanaacactctctatgcaacaacagccacaaatccca	2614
Db	1178	TCCGATTTTGCAGAGACTTTTAAACAAAACCACTCTATGCAACAACAGCCACCAAAATCCCA	1237
OY	2615	cccttctcatctctcagatctggacaagaatgtgtgtttagtctccttgacaacatctgga	2674
Db	1238	CCTTCTCATTTCTCTGCACAAATTGGACAAAGAAATTTGTGTTTAACTCTTGACAAATGTGAA	1297
OY	2675	gcattgagcagctgylaaccgcgctccaatgcagtgaaatcaagcagcgcaagcttctcgtctgc	2734
Db	1298	GCATGGCGACACTGGTAACCGGCTCAATTCGACTGAATCAACACAGGCAAGCTTTTCCGTGTC	1357
OY	2735	agacaagttagagctctggggtcctctgggtctgggaatgtgtgacatttgaacagtctgcgccatgac	2794
Db	1358	AGACAGTTGACCTGGGCTCTCGGTTGGGATGGTGACATTTTGAACAGTGTCTGCCCATGTGAC	1417

QY	2795	aaatbtgaactatatacagaataaagaatgagcaagttagacagggagacacactgtccaaagaattac	2854
Db	1418	AAAGTGAACCTCATATACAGATATAAGATGCGATGACAGGACACACTCCGCAAAAGATTAC	1477
QY	2855	cttgacagcaagcttccagggaggagacgtlccactctgcagcagcgggtcttcgactcgattactgata	2914
Db	1478	CTTGACGACGCTTCAGSAGGAGCGTCATCGCAGCGGCTTGATCGGCATTACT	1533
QY	2915	tgtggcaacatttgcctglttlltccatgacacacagcagttaatggggaatgtgcagaagaa	2974
Db	1534	-----	1533
QY	2975	atccaaattgycctctctgycctgtcagcttaagttaagaagaatatccaaactgagt	3034
Db	1534	-----GTGATTTGGAAGAAATATCCAACTGATG	1561
QY	3035	gacttgaatatgtgtctgtctgcagatagtgagggaagaacaacataaagtgtgtgttctaag	3094
Db	1562	GATCTGAATTTGTGCTTCCTCAGCGAGTGGGAAACACTATTAAGTGGGTCTTTAACG	1621
QY	3095	aggttcaacaagaagtgtgtgtccatctccacaacagtgcttttgggcctctcagctcaag	3154
Db	1622	AGGTCAAAACAAGTGGTGCCATATCTCACACAGTCCGCTTTGGGGCCCTCTGCAGCTCAAG	1681
QY	3155	aactagaggaagctgtgccaaaaatgacaggaaggtttaagaacataatgtcttcagatcaagtlc	3214
Db	1682	AACTAGAGAGACTGTCCAAATATGACAGAGAGGTTTACAGACATATGTCTTCAGATCAAGTTC	1741
QY	3215	agaaacaatgacctatattgactgttlltggggcccttcatcaagaaatgaagtgctgtctc	3274
Db	1742	AGAAACAAGGCTCTATTTAGTGTGTTGGGGCCCTTTTCATACAGAAATGGAAGTGTCTCTC	1801
QY	3275	agagctccatccagactgtgagagtaagaagattaaacctccagaacagccagtgatgtaatg	3334
Db	1802	AGCGCTCATCTCACCTTGAGAGATGAAGGATTAACCTCCAGAAACGCCAGTGAGTGAATG	1861
QY	3335	gcaacagtatctgtgtgacagcacccgttgggaagaagcaacttgttcttatacctcggaaaca	3394
Db	1862	GCAACAAGTATGTGTGACAGCACCGGTGGAAAGACACTTGTGTTCTTTACCTTGACCTGGAACA	1921
QY	3395	tgcagcctccccaatctctctctgtgattcccaagtgcagaagaagaagtgtgctttag	3454
Db	1922	CGCAGCCTCCCAATGCTTCTCTGGGATGCCAGTGTGACAGAAAGCAAGTGGCTTTGTAG	1981
QY	3455	tggacaaaaaacccaataatgtgcttactcccaatctcccaaggaatgtctaaagttgtgcact	3514
Db	1982	TGGACAAAAAACCCAAATATGGCTTACTCTCCAAATCCCAAGCATTTGCTTAAGGTGGCACTT	2041
QY	3515	ggaataaagctctgcagcagaagcttcaaaaaccttgcacctgtacgttccagtcctgtgt	3574
Db	2042	GGAATATACGTTCTCAAGGCATCACAACCTTGTGACCTTGACTGTACGTGCCGTGCT	2101
QY	3575	ccaattgttacctgtcctccaattacatgacttgcacaaaagcaacaagaagacacacagcaaat	3634
Db	2102	CCAATGCTACCTCTGCTCCATATTACAGTGACTTCCAAAACGAAACAAGACACACACCAATAT	2161
QY	3635	tcccagacccctctgtgtagttaatgtcaaatatllcccaagaagcctccccaattctcagag	3694
Db	2162	TCCCCAGCCCTCTGTGTGTTATGTGCAAAATATTCGCCAAGAGCCCTCCCAATTTCTCAAGG	2221
QY	3695	ccaagtgtcaacagcccttaattgaatcagtgatgtgaanaaacaatlaaccttgcgaactacg	3754
Db	2222	CCAGTGTACACGCCCTTATTTGAATCACTGAATGTGAATGAAAAACAATTACCTTGGAAACTACGG	2281
QY	3755	ataatbggaacagtgctgtatgtactaagaatgaoggttctactcoaagtatttccaca	3814
Db	2282	ATAATGGAGCGAGTGTGATGCTACTTAAGATGACCGTCTTACTACCAAGATATTTCACAA	2341
QY	3815	cttttgacaagaatgtaagaatacaggttaaaaatgtgggctcttgggaagagttaacag	3874
Db	2342	CTTTATGACCAATATGGTATGATACAGTATAAATGTGGGGCTCTGGGAGAGATTAAACGAG	2401
QY	3875	ccaagcggaggtatataccccaagabaagtgtgaagcaatgttataactgtgtctgtaattgga	3934

Db	2402	CCAGCAGGAGAGTATACCCAGCAGAGTGAGCAGCTGTAATACCTGGCTGGATTGAGA	2461
Qy	3935	atgatgaaatcaatggaatccccaagaacctgaaataataaagatgatygttcaaca	3994
Db	2462	ATGATGAATTCATATGAGATCCACCAAGACCTGAAATTAATMAAGATGATGTTCCACACA	2521
Qy	3995	agcaagtgtgttcaagcagaacatccctcgggaagctaatltgtgtcttgatgttccaa	4054
Db	2522	AGCAAGTGTGTTTAGCAGCAAACTCTCGGAGGCTCATTTTGTGTGCTTGATGTCCCAA	2581
Qy	4055	atgtctccataacctgattctctccacctggtccaaatcacccgaacctgaagcggaatc	4114
Db	2582	ATGCTCCATACCTGATGATCTCTCCACACTGGSCAAATACACGACTGGAAGCGGAATTTC	2641
Qy	4115	acgggggcgaagcttaattatcttgcttggaaagctctctggggatgatatgacatygaa	4174
Db	2642	ACGGGGGCAAGCTCATTAATATCTGACTGTGAGACAGTCTCTGGGGATGATTATGACCATGAA	2701
Qy	4175	cagctcacaaatatacatctcgaataagtaacgaatattcttgatctcagagacaagttca	4234
Db	2702	CAGCTCACAAATATATCATCTTGATGATAGTACAAATATTCTTGATCTCAGACACAGTTCA	2761
Qy	4235	atgaatctctcaagtgaataactacatgctctcatccatccaagaagccaaactctgaagaa	4294
Db	2762	ATGAATCTCTTCAAGTGAATCTACTGCTCTCATCCCAAGGAAGCCAACTCTGAGGAAG	2821
Qy	4295	tcttttgtttaaacccaagaacaattacttttgaanaatgagacagatctttcaigtcta	4354
Db	2822	TCTTTTGTTTAAACCAAGAAACATTTACTTTTGAAATATGGCACAGATCTTTTCAATTGCTA	2881
Qy	4355	ttcagagctgttgataaagtcgatcttgaaatcagaataatccaaacttgacagaglatcct	4414
Db	2882	TTCAAGCGTGTGTAATAGATCGATCTGAAATCGAATAATCCAAATCATGACAGATATCTT	2941
Qy	4415	tgttattatccccaagaagctccgccaagaagacacttagtctctgataagaagcttgcctct	4474
Db	2942	TGTTTATTCTCCACAGACTCCGCCAAGAGAACCTAGTCTCGATGGAAGACCTGTGCTCTT	3001
Qy	4475	gtcccaatatcatatcaacagacacattcctgtgcatcacatlttaaaatatgtgga	4534
Db	3002	GTCCCTAATATTCATATATCAACAGACCAATTCCTGTGCATTTCACATTTTAAATATATGGA	3061
Qy	4535	agtgtatagagggaactgcgaagctgtgtcaatagcctag	4569
Db	3062	AGTGATAGGAGAACTGCAGCTGTCAATAGGCTTAG	3096
RESULT	3		
LOCUS	AF127036		
DEFINITION	AF127036	2826 bp	mrna
ACCESSION	AF127036		
VERSION	AF127036		
KEYWORDS	AF127036.1	GI:4585468	
SOURCE			
ORGANISM			
			human.
			Homo sapiens
			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE			1 (bases 1 to 2826)
AUTHORS			Aguel,M., Vernat,T. and Colouscou,J.M.
TITLE			Identification of three novel members of the calcium-dependent chloride channel (CacC) family predominantly expressed in the digestive tract and trachea
JOURNAL			FEBS Lett. 455 (3), 295-301 (1999)
MEDLINE			99364503
PUBMED			10437792
REFERENCE			2 (bases 1 to 2826)
AUTHORS			Aguel,M. and Colouscou,J.-M.
TITLE			Direct Submission
JOURNAL			Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des Carrières, Rueil-Malmaison 92500, France

FEATURES		Location/Qualifiers
source		1..2826 /organism="Homo sapiens" /db_xref="taxon:9606" /tissue_type="small intestine; colon"
gene		1..2826 /gene="CACCL1"
CDS		5..2749 /gene="CACCL1" /note="bovine epithelial chloride channel homolog" /codon_start=1 /product="calcium-activated chloride channel protein 1" /protein_id="AAD25487.1" /db_xref="GI:4585469"
BASE COUNT	875 a 623 c 632 g 696 t	
ORIGIN		"
Query Match	54.1%: Score 2473.6; DB 9; Length 2826;	
Best Local Similarity	96.3%: Pred. No. 0;	
Matches 2582; Conservative	0; Mismatches 4; Indels 96; Gaps 1;	
QY 1888	aagagcatgcygacccaagcatctctgcatctgcttgctgaagcttacaggaagcgattttat	1947
DB 164	AAGGACATGGTGACCCAGGACATCTCTGATCTGTTTGAAGCTACAGGAACGATTTTAT	223
QY 1948	ttcaaaaatgltgccatttgcatttgcctcgaacaatggaacaaagcctgactatgtgaga	2007
DB 224	TTCAAAATGTTGGCATTGTTGATTCCTGAACATGGAAGCAAAAGGCTGACTATGTGGA	283
QY 2008	ccaaaacttgagacctataaaaaatgctgatagttcttgcttgctgagttctactcctccaggt	2067
DB 284	CCAAAACCTTGAGACCTACAAAAAATGCTGATTTTCTGGTTGCTGAGTCTACCTCCCAAGT	343
QY 2068	aatgataccctcaacgcagacatgagcaacgttgagagaagaggtgaaagatccac	2127
DB 344	AATGATGACCCCTACACGTACGACAGATGGGCAACTGTGAGGAAGAAAGGTGAAGAGATCCAC	403
QY 2128	ctcactccgtatcttcatttcagagaaaaaagttaagctgaatactggaaccacaagtgtaggca	2187
DB 404	CTCACTCTCTGATTTTCATTCAGGAAAAAAGTTAGCTGAATATGAGCCACAAGTAGGGCA	463
QY 2188	tttgccatgagtgaggctccatctacagatggggagatatttgcagagttacaataatgatgag	2247
DB 464	TTTGTCCATGAGTGGGCTCATCTACGATGGGGAGATTATTGACGAGTCAATATGATGAG	523
QY 2248	aaatctcctattccaatggaagaataacaacgagtaagatcttaagcaggtatctactggt	2307
DB 524	AAATTCCTATTATCAATGAGGAAGATACACAGCAGTAAGATGTTATGACGAGGTATTACTGCT	583
QY 2308	acaaatgtagtaaaagatgtgcagggagcagcgtttacacacaaagaatgcacattcaat	2367
DB 584	ACAAATGTAGTAAGGAAGTGTACAGGAGCGCAGCTGTATACACAAAGATGACATTCAT	643
QY 2368	aaatgaagaagactctatgaaaaagatgtgagtttgttccaatccgcagagag	2427

Db	644	AAAGTACAGGACTCTATGAAAAAGATGTGAGTTGTTCTCCAAATCCCGCACAGCGAG	703
Qy	2428	aaggtctctaatagtttgcaacaatgttgattcttaagttaattctgtacaagaaca	2487
Db	704	AAGGCTTCTATATGTTTGCAACACATGTTGATTCTTAGTTGAATTCGTACAGAAACAA	763
Qy	2488	aaccacaacaagaagcttccaacaagaacaatccaatgaatgcactccgaagacata	2547
Db	764	AACCACAACAAGAGCTCCAAACAGCAAAATCAAAAATGCAATCTCCGAAAGACATAGG	823
Qy	2548	gaagtatccgttatcttgagactttaagaacaacccctatagacaacagcacca	2607
Db	824	GAAGTGAATCCGTGATCTTGAGGACTTTAAAGAAAACCACTCCTATGACACACACCAACA	883
Qy	2608	aatcccaaccttcacatgtctgcgaatgtggaacaagaatttggtgttaagtccttgaaaa	2667
Db	884	AATCCCACTTCTCATATTGCTGCGAATTTGACAAAGAAATTTGTTTATGCTTTGACAAA	943
Qy	2668	tcgtgaagcaltgagcagcttglaaccgcctcaatcgactgaatcaagaagcgcaagctttc	2727
Db	944	TCGTGAAGCATGGCGACTGGTAACCGGCTCAATGACTGAATCAAGCAGGCGACTTTTTC	1003
Qy	2728	ctgctgagaagatgtgctgctgggttctggtttgggatggtgacatttgaaagtgcgc	2787
Db	1004	CTGCTGAGACAGTTGAGCTGGGGTCTGGGTTGGATGGTGAATTTGACACATGCTGCC	1063
Qy	2788	catgtacaatatgaactacatacagataaacagttgcagttgacaggaacacatcgccaaa	2847
Db	1064	CATTTACAAAGTGAACCTATACAGATTAACAGTGGCAGTGAACAGGACACACTGCGCCAAA	1123
Qy	2848	agattacctgcagcagcttcaagaaggagcttccatctgcagcgggcttcgaatgcgcaatt	2907
Db	1124	AGATTACTCTGCAGCAGCTTCAGAGGAGCGTCATCTGCACGCGGCTTCGATCGGCATTT	1183
Qy	2908	actgatattgaggcaactttgctgtttccatgacacacagcagttatggggagtgca	2967
Db	1184	ACT-----	1186
Qy	2968	caagaanaatccaatltggcctctctgctgcagcttaagttaagaanaatatcca	3027
Db	1187	-----GTCATTAGGAAGAAATATCCAA	1207
Qy	3028	actgtatgatctgaattgtgtgtcgtgaacgaatggggaagaacaacataaagtgggtgc	3087
Db	1208	ACTGATGATCTGAATTGTTGCTCTGACGAGTGGGAAGACACACATTAAGTGGGTGC	1267
Qy	3088	tttaacagagtcacaacaagtgtgtgcataatccacaacagtcgctttggggcctctgca	3147
Db	1268	TTTAAACAGAGTCAAAAGTGTGTCATCATCCACACAGTGGGCTTGGGGCTCTGCA	1327
Qy	3148	gtccaagaactagaggagctgttccaatatgacagaggtttacagacataatgtctcaagat	3207
Db	1328	GCTTAACACTAGAGAGCTGTCCAAATGACAGAGGTTTACAGACATATGCTTCAGAT	1387
Qy	3208	caagttcagaacaatggcctcaatgtatgtctttggggcctttcatcagaagaattgagct	3267
Db	1388	CAAGTTGAGAAACATGCGCTCATTTGATGCTTTGGGCGCTTTCATCAGGAAATGAGACT	1447
Qy	3268	gttccttaagcgtccatccagcagcttgaaagtaagagtaaacccctcagaagaagcaatg	3327
Db	1448	GTCTCTAGCGCTCCATCCAGCTTGAGATTAAGGATTAAACCTTCAGAAACAGCAGTGG	1507
Qy	3328	atgaatggcaagatgcgttggaacagcacgctgggaagaagcaactgttctctataacc	3387
Db	1508	ATGAATGGCACAGTATCGTGGACAGCCTGGGAAAGGACACTTGTCTTATATACCC	1567
Qy	3388	tggaacaatgagcctccccaatccttctbgygaatcccaatgggacagaagcaagtggc	3447
Db	1568	TGGACACAGCAGCTCCCAAAATCTCTCTGGGATCCCAATGAGACAGAAACAGAGTGGC	1627
Qy	3448	tttgtatggaacaaaacacaaaatggtcctactcctccaatccaggcatgtcctaagttc	3507
Db	1628	TTTGTATGAGACAAAACACCAAAATGCGCTTACCTCCAAATCCAGGCAATTGCTAAGGTT	1687

---

Qy	3508	ggcacttggaaatacaagtcctgcaagcaagcttcaacaacttgaccctgactgtcaagctc	3567
Db	1688	GGCACTTGGAAATACACTCTGCAAGACAGCTCACAAACCTTGACCTGCATGTCAGCTCC	1747
Qy	3568	ctgtgcctcaatgtctaccctgtcctccaattacagtgacttccaaaacgaagaagcacc	3627
Db	1748	CGTGCTGTCAATGCTTACCTGCTCCCAATTAACAGTGCATCTCCAAAAGAAAGAGACACC	1807
Qy	3628	agaaatctcccaagcctctggtagttaatgcaaatatctggcaagaagcctcccaatt	3687
Db	1808	AGCAAAATCCCAAGCTCTGTGATGTTATGCAAAATTTGCCAAGGAGGCTCCCAATT	1867
Qy	3688	ctcagggccagtgatcaagccctgattgaaatcagttgaatgaaanaacagttacttgaa	3747
Db	1868	CTCAGGCGCAGTGTACAGCCCTGATTTGATTCAGTGAATGGAATAACAGTTACTCTGAA	1927
Qy	3748	ctactgataatggaagcaggtgtcgtatgtactaaggaatgaaggtgtctactcaaggtat	3807
Db	1928	CTACTGATTAATGAGCAGGCTGTGATGCTTAAAGTGAATGATGACGCTGCTCAAGGTAT	1987
Qy	3808	ttcaacaactatgacacgaatgtgtatagatgtataaagtgcgggctctgggaaggttc	3867
Db	1988	TTCAACAATTATGACAGCAATGGTAGATACAGTGTAAAGTGCAGGCTGTGGAGAGTT	2047
Qy	3868	aacgcagcagaagagatgatacccccaagcagatgggaagcactgtatacactgtgc	3927
Db	2048	AAGCAGCCAGAGGAGAGATATACCCAGAGAGTGGACAGCTGTACATACCTGGCTGG	2107
Qy	3928	attgagaatgtgaataacaaatggaatccaaagacctgaatataaagatgatgtc	3987
Db	2108	ATTGAGATATGAAATACATAGAAATCCAAAGACCTGAAATTAATTAAGATGATGT	2167
Qy	3988	caacacaagaagaagtgttctcagacagacatccctcgggaagctaatgttggctctgat	4047
Db	2168	CACACAAAGAGTGTGTTTTCAGCAGAACATCTCTGGAGGCTCATTTGTGGCTTCGAT	2227
Qy	4048	gtcccaaatgtcccatcactgtactcttcccaacttggccaatatcagcacctgaagcg	4107
Db	2228	GTCCCAATGTCTCCATACCTGATCTCTTCCACCTGGCCAAATCAGCAGCTGAAAGCGC	2287
Qy	4108	gaaatcagggggagcttccatataatctgaacttggaacgctctcgggagatgattatgac	4167
Db	2288	GAAATTCACGGGGCAGCTTCATTAATCTGACTTGGACACTCTCTGGGATGATTATGAC	2347
Qy	4168	catggaacagctcaagaatatalcatccgaataagtaagaatattcttcttctccaagac	4227
Db	2348	CATGGAACAGCTCACAGTATATTCGANTATGACAGTATTCCTTGATCTCAGAGAC	2407
Qy	4228	aagttcaatgaatccttctcaagtlgaatlaactgtctctcatcccaagaagaaccaactc	4287
Db	2408	AAGTTCAATGAATCTCTCAAGTAATACTACTCTCTCATCCCAAGAACCAACTCT	2467
Qy	4288	gaggaaagctcttttgtttaaaccaagaanaacatctctttgaanaatgcaagaatcttcc	4347
Db	2468	GAGGAAGTCTTTTGTGTTAAACGAAACATTAATCTTTGAAATGGCACAGATCTTTTC	2527
Qy	4348	atgactatcagagctgttgaatgaatgcgactggaatcagaataatccaacatgtcaga	4407
Db	2528	ATTGCTATTAGGCTGTTGATTAAGTGTGATCTGAATTCAGAAATATCCAAATTTGCCAGA	2587
Qy	4408	gtactcttgttlatcttccacagactcgcgcagagaacacactagtcctgtatgaaagctc	4467
Db	2588	GTATCTTTGTTTATTCCTCCACAGACTCCGCGAGAGACACTAGTCTGATGAACGCTCT	2647
Qy	4468	gtcccttgctcctaatalatcatatcacagacaacatctcgtgcatatcaatttaanaatc	4527
Db	2648	GCTCTTGTCTTATATATTCATATCAACAGCACCAATTCCTGGCATTCACATTTAAAAATT	2707
Qy	4528	atgtggaagtgtatgagaagaatcagctgtcaatagcctgac	4569
Db	2708	ATGTGAAGTGAATGAGAGAACTGCAGCTGTCAATTAAGCTTAC	2749



```

RESULT 4
AK024970
LOCUS AK024970 2022 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLN21317 fls, clone COL02275, highly similar to
AP127036 Homo sapiens calcium-activated chloride channel protein 1
(CACCI) mRNA.
ACCESSION AK024970
VERSION AK024970.1 GI:10437397
KEYWORDS oligo capping; fls (full insert sequence).
SOURCE Homo sapiens colon cDNA to mRNA, clone_lib:COL clone:COL02275.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE
AUTHORS Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
JOURNAL Unpublished (2000)
REFERENCE 2 (bases 1 to 2022)
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction: 5'-6' and one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
SOURCE 1. 2022
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL02275"
/clone_lib="COL"
/lssue_type="colon"
/note="Cloning vector pME18SFL3"
misc_feature 1..2022
/note="highly similar to AP127036 Homo sapiens
calcium-activated chloride channel protein 1 (CACCI) mRNA"
BASE COUNT 612 a 472 c 453 g 485 t
ORIGIN
Query Match 40.4%; Score 1847.2; DB 9; Length 2022;
Best Local Similarity 95.2%; Pred. No. 0;
Matches 195; Conservative 0; Mismatches 3; Indels 96; Gaps 1;

```

```

Db 241 GGGTTGGATGGTGCATTGTGACAGTGCCTGCCCATGTACAAAGTCACTCAATGATTA 300
Oy 2816 acaagtgcagatgcagaggaacacactcgcgaagaagatacctgcagcagcttcagaaggga 2875
Db 301 ACAGTGGCAGTGCACAGGAGACACACTCGCCAAAAGATTAACCTGCACAGCTTCAGAGGGA 360
Oy 2876 cgtccatctgcagcggggttcgatgcattacgatatatgtagcaaatcttgctgttt 2935
Db 361 GCTCATCTGCAGCGCGGCTTCGATCGCATTTACT- 395
Oy 2936 tccatgcacacagcagttatgtaggagtcgcagaagaataatccaatctggcctctcgg 2995
Db 396 ----- 395
Oy 2996 cctgcagcttgatgtagtaggaagaatatccactgatgatctgaatctgctgtcga 3055
Db 396 -----GTGATTTAGGAAGAAATATCCAACTGATGATCTGAATTTGCTGCTGA 444
Oy 3056 cggatggggaagacacacactataagtggtgtcttaacgaggtccaacaagtggtcca 3115
Db 445 CGGATGGGGAAGACACACTATTAAGTGGGTCTTTAAACGAGGTCAAAACAAAGTGTGCA 504
Oy 3116 tcatccacacagtcgcttggtggccctgcagctcaagaacatagaagagctgtccaaa 3175
Db 505 TCATCCACACAGTGCCTTTGGGGCCCTCTGCAGCTCAAGAACTAGAGAGCTGTCCAAA 564
Oy 3176 tgcagagaggtttacagacatatgcttcagatcaagttcagaacaaatggcctcattatg 3235
Db 565 TGACAGGAGGTTTACAGACATATGCTTCAGATCAAGTTCAGAACAAATGGCCATTTGATG 624
Oy 3236 ctttgaggcccttcacaggaatagagcgtgtctcagagccctccagcttgaga 3295
Db 625 CTTTGGGCGCCCTTTCATCAGGAAGAAATGAGCTGTCTCAGGCTCCATCCAGCTTGAGA 684
Oy 3296 gtaagggatlaaccctccagaacacagccagttgatlaagtcagtcgtgcagca 3355
Db 685 GTAAGGATTTAACCTCCAGAAACAGCCAGTGATGAATGGCAGCTGATCGTGACAGGA 744
Oy 3356 ccgtgggaaggagaacctgtttcttctacactcgtgcagaaatgcagcctcccaatccttc 3415
Db 745 CCGTGGGAAGGACACTTTGTTTATTCACCTGGACAAACGACGCTCCCAAAATCCTTC 804
Oy 3416 tctgggacccagtgagaagaagcaggtgtgtgttagtggacaaaacacaaatg 3475
Db 805 TCTGGGATCCAGTGGACAGAGCAAGGTGGCTTGTGTGGACAAAACACCAAAATG 864
Oy 3476 cctactcctcaaatcccaaggcattgctlaaggttgcacttggaaatacagcttcgaagca 3535
Db 865 CTTACTCCCAATCCCAAGSCATTCCTAAGGTTGGCAGCTTGAATTAACAGTTCGACAGCA 924
Oy 3536 gctcaacaaccttgaccctgcactgcacgtccgtgtgcctcaatgctcctccaa 3595
Db 925 GCTCAAAACCTTGACCTTGACTGTACGTCGCCGCTCCAAATCCTACCCGCTCCAA 984
Oy 3596 ttacagtgacttccaaaacgaagcagaagacacgaagaattcccgagcctcgtgagttt 3655
Db 985 TTACAGTCACTTCCAAACGAACAGACAGACACCAAGAAATCCCGAGCCCTGTGGATTT 1044
Oy 3656 atgcacaatatccgcaagagagcctcccaatctcaaggccagttgcaagccctgattg 3715
Db 1045 ATGCAAAATATTCGCCAAGAGAGCTCCCAATTCAGAGGCGAGTGTCAAGCCCTGATTG 1104
Oy 3716 aatcagtgataggaagaacaggttaccttggaactcgtgataatgtagcaggtgtcgtatg 3775
Db 1105 AATCAGTGAATTCGAAGAAACAGTTTACTTGGAACACTGATGAATAGGACAGTGTGATG 1164
Oy 3776 ctactaagatgagcgtgttactcaaggtatttcaacaattatgacagaaatgtaat 3835
Db 1165 CTACTAAGATGACGCTGTACTCTCAAGGTATTTTACAACTTATGACAGAAATGTTAGAT 1224
Oy 3836 aacagtlaaaggtgcggcctctggaggagttlaagcagcagcagagaggtgataccccc 3895

```

DB	1225	ACAGSTTAAAGGCGGGCTGTGGAGGAGTTAAACCCACCAAGAGAGATGATNCCC	1284
QY	3896	agcagaagtggagcactgtatcatcactgtgccttgatctgagaatgatgaatcagaatgc	3955
Db	1285	ACCAAGTGTGACACACTGTACATACCGCTGGATTGAGATGATGAAATACATGGAATTC	1344
QY	3956	caccagaacctgaatlaaagaatgatgcttcaacacagaagaagtgtgttcaagcaga	4015
Db	1345	CACCAAGACCTGAATTAATTAAGATGATGTTCAACACAAAGCAAGTGTGTTCCAGCAGNA	1404
QY	4016	catcctcgggaagctcatttgtgtgtcttgatgltcccaatgctcccaatacctgactct	4075
Db	1405	CATCCCTGGGAGGCTCATTTGTGGCTTCTGATGATGCCAATGCTCCCAATCCGTGATCTCT	1464
QY	4076	tccacacctggcaaatcaccagcactgaaagcggaattcaacggggcgacttccatcattac	4135
Db	1465	TCCCACTTGCCCAATCACCGACCTGAAAGCGGAATTCACGGGGGCACTGCATTAATTC	1524
QY	4136	tgacttggaagctcctctggggaatgaltatgacatlgaaacagctcacaaatlatcatc	4195
Db	1525	TGACTTGGACAGCTCCTGGGATGATTAAGACCATGGAACAGCTCACAGTATATATTC	1584
QY	4196	gaataagtcacagtattcttgatctcagaagacaagttaaagaatctcttcaagtgata	4255
Db	1585	GAATAAGTACAAAGTATCTTGATCTCAGAGACAAAGTCAATGAATCTCTTCAAGTCAATA	1644
QY	4256	ctacgtcctcatcccaagaagacccaactcctgagaagctcttttgtttaaaccgaa	4315
Db	1645	CTACGCTCTCATCCCAAGGAAGCAACCTCGAGGAATCTTTTGTTTAAACCAAGAA	1704
QY	4316	acattactttgaaaatlgcagacagatctttcatgtctatccaagctgttgaagtcg	4375
Db	1705	ACATTACTTTTAAATATGACACAGATCTTTTATGCTATTCAGGCTGTGTAAGAGTGC	1764
QY	4376	atctgaatcagaataatcccaactgcacagatgatactttgttatactcccaagactc	4435
Db	1765	ATCTGAATACAAATATTCACATTCGACGAGTATCTTTGTTTATTCCTCCACAGACTC	1824
QY	4436	cgcagagaacactatgctctgaatgaacgctcgtcctctgttccataatcatatcaaca	4495
Db	1825	CGCCAGAGACACCTGTGCTCTGATGAAGCTGCTCCTGCTTAATATTCATATCAACA	1884
QY	4496	gacacattcctggacattccattttaaataatgatgtgaaagagatagagaacatgcgc	4555
Db	1885	GCACCAATCTCTGGCATTCACATTTTAAATATATGTGGAGATGAGATGAGAACTGCACG	1944
QY	4556	tgatcaatagcctag 4569	
Db	1945	TGTCAATAGCCTAG 1958	
RESULT	5		
AF095584			
LOCUS	AF095584	3079 bp	mRNA
DEFINITION	Sus scrofa epithelial chloride channel protein (ABCC) mRNA,		
ACCESSION	AF095584		
VERSION	AF095584.1	GI:6002645	
KEYWORDS			
SOURCE			
ORGANISM			
	Pig.		
	Sus scrofa		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.		
REFERENCE			
AUTHORS	Gaspar,K.J., Racette,K.J., Gordon,J.R., Loewen,M.E. and		
	Forsyth,G.W.		
TITLE	Cloning a chloride conductance mediator from the apical membrane of		
JOURNAL	porcine ileal enterocytes		
MEDLINE	Physiol. Genomics (Online) 3 (2), 101-111 (2000)		
PUBMED	20473747		
	11015605		
REFERENCE	2 (bases 1 to 3079)		
AUTHORS	Gaspar,K.J., Gabriel,S.E., Racette,K.J. and Forsyth,G.W.		

TITLE		Direct Submission
JOURNAL		Submitted (28-SEP-1998) Veterinary Physiological Sciences,
		University of Saskatchewan, 52 Campus Drive, Saskatoon, SK S7N 5B4,
		Canada
FEATURES	Location/Qualifiers	
source	1..3079	
	/organism="Sus scrofa"	
	/db_xref="taxon:9823"	
	/tissue_type="ileal mucosa"	
	1..3079	
gene	/gene="AECC"	
	133..2886	
CDS	/gene="AECC"	
	/function="involved in exocrine chloride conductance"	
	/codon_start=1	
	/product="epithelial chloride channel protein"	
	/protein_id="AF00077.1"	
	/db_xref="GI:6002646"	
	/translation="MGFSRSLFTLVHLLEGAQSNLSIQNGNGEYGIYAIADPNVP EDELINIKDMYTKAPTEFEATGEKEKRYFNPAIIPAMKRPVEVKRLKFNKKA DVVTEEPENKPNPSPYEMONGNCEKEKLYPMDIPAGKGLYVGPGGVPEHNAH LRNVPENYNECKEFTLSNKKEDPYISAIIRGTNULPQCGSCVTKPRRARVITGL FQKCEFTIPDQSEKASIMEAOSIDIVPECEKNNKAPADQNKCVLRSTWEY QDSDEKFKTPTPTTPPPAPFSLQIQIVCLVDKSGSMVGRLKRLNQKLEFL LQYEOGAGWGVAFDSAAVYKSELVQINSAEADLARSIPYASGGSIGSLRSA FTVKKYKPDGSEITVLTDGEDNTIACPEYKONALIHIVALGSAKLELEISO MTGGLIYASDQANENGLIDAFGLSSGNSASORSITOLEGKGLTTONENMGATVY DYGKRLPLFLITLERKFTSPIPEGVSGRSODSPLYGKHNKATFOVGTAKVGM KYSIQASOTLITLVSSRRSATLPVTVISKMKDGRKSPVAVYTKIHQGLPL RAKATYILESENKTVTLEILDNAGADQIRVNDISRYTAPADANGRSFKWALGG VNPERRAPLEMSGAMYIRGMVINGELIKMNPBRNDIKDLOQKOFCSFTASGSFV ASDVKSPIDLEPPCKITDLAGIQDNLINLTPAFGDDYDHGRADRYIIRISNI LDLDEKFNDSVQVMTDLIRKANSESEVFEKPEGIFPTGTDLFLAVQAVDNTLKS EINSIAQVSLFLPEAPERPPEPAPSLCPLEIYVNSTIPGIIHKIMKMWIGELQL SIA"	
BASE COUNT	892 a 758 c 755 g 674 t	
ORIGIN		
Query Match	35.8%; Score 1637.8; DB 4; Length 3079;	
Best Local Similarity	77.4%; Pred. No. 0;	
Matches 2084; Conservative	0; Mismatches 502; Indels 105; Gaps 4;	
OY 1888	aagacacgctgacccagacacgacacgctctgcatctgcttgaagctacagaagaacgatttat 1947	
Db 292	AAGGATATGGTGACCAAAAGCATCTCCATATCTGTTGAAGCTACGAGAAAAGATTCAT 351	
OY 1948	ttcaaaaatgttcacatttgaatcttgaacccgaagaacagagctgacatagtgaaga 2007	
Db 352	TTCAAAATGTGGCATTTTTATCTCTGCAAGGTGGGAAGCAAAACCGAATATGTGAAA 411	
OY 2008	ccaaaacttgagacctacacaaaatgctgaatgcttgcgttgctgagcttactctccagat 2067	
Db 412	CCCAAACTTGAGACCTCAAAAATATGCTGATGTTGTGGTTACTGAACCAAAATCCCTCGAAA 471	
OY 2068	aatgatgaacctctacacgagacagatgagacacgctgtagaagaaggttgaagaattccac 2127	
Db 472	AATGATGGCCCTTATACGAGACAGATGGGAAACGTGCGACAAAAGGCTGAAAAAATTTAT 531	
OY 2128	ctcacctctgaattcatctgacgagaaaaaagttaagctgaattgaacccaagaagttagca 2187	
Db 532	TTCACTCTCTGACTTCGTAGACGAGAAAAAGGTGCTTCAATATGACCAACAAGGAAGGTA 591	
OY 2188	tttgatcaatgagtggtcatalctacagatggggaglatlttgaagatgaacaatatgatgag 2247	
Db 592	TTTTGTCACGATAGGGCTCACTTCGATGGGGCGCTGTTTATATGATGATCAACAAACGACACAG 651	
OY 2248	aaattctactatccaatgtgaagaataacaaagcagtaagaatgttcaagaagttattacgct 2307	
Db 652	AAGTTTACTTATCCAAACAAAAGAAACAAACCACTGATATGTTTGCGAGCCATCCGGGT 711	
OY 2308	acaatgtagtaagaaggtgtcaggaagagcagctgttaccacaaagaatgcacattcat 2367	

Db	712	ACAATGTACTACCAACAATGCCAAGGACGACGCTGTCGTACCAAAACCGTGGCAGACAG	771
QY	2368	aagtaaacgaagcctcta tgaanaagga tctgaat tttgtcttccaa tccgcgcagacgag	2427
Db	772	AGAGTAACGGGGCGTTCGCAAGAGAGATGTGATTCATACCGATCCAGACGAGCGAG	831
QY	2428	aagcctctataat tgttttgacacaacat tgttata tctatagttgaat tctgtacagaca	2487
Db	832	AAGGCCTCTATAATGTTTGACACAGACGATGATCTGTGGTGAATTCGTAAAGAAAA	891
QY	2488	aaccacaacaaagagcttccaaaacagcaaat ccaanaa tgcnaa tctccgaagcaca tgg	2547
Db	892	AACCCACATTAAGAAAGCTCCAAACGACCAAAACCAAAAAATGCAATCTCGAAGCAATGG	951
QY	2548	gaagtatcctgtat tcttgagagctttaaanaaacat cctctatga caaacagacacca	2607
Db	952	GAGTGATCCCAAGTTCTGTGAAGACTTTAAACAAACCACTCCCATGACAACAGCCACC	1011
QY	2608	aatccacatctcatal tgcctgcagat tgcacaaagaat tgttgttagt cctctgacaa	2667
Db	1012	GCGCCACCTTCTACGTCTGCAGATGCGACAGAAATGCTGTACTACTCTGCAGAA	1071
QY	2668	tctggaagcagtgagcactgtgttaacccgcttaa tctgcactgaatccaagcagccagctttc	2727
Db	1072	TCTGGAACCAAGACGGTGTGTGGCGCTTTAAACGACTAAATCAAGCAGCAAACTCTTC	1133
QY	2728	ctgcctgcagacagcttgagctgcggagctctgtgttggttgagatgtacat tttgaatgtctgc	2787
Db	1132	CTTCTGCGAGACAGTGGACCAAGGGGCGCTGGGTGTGGATGGTGCAATTTGACAGCGCTGC	1191
QY	2788	catgtacacaaa tgaactcatacagataaacaatgtgcagtgcagtgagcagggacacatctgc	2847
Db	1192	TATGTAAAAATGAACTCGTCACGATTAACAGCGCGCGGAAAGAGACCGCTGTCCGCG	1251
QY	2848	agatatactgtcagacagcttcaaggaaggagctccatctgcagcggccttgatcgcacatt	2907
Db	1252	AGCTTATCCCAACAGACGAGCTCGGGGAGCAACCTCATCTGCTGGGGCTTGTGCTCCCATTT	1311
QY	2908	actgatatgtgcacaat tgcctgttctccatgcacaaacagcagttatgggagtgga	2967
Db	1312	ACT-----	1314
QY	2968	caagaaaatccaat tiggcctctctg cctcgtcagccttagtgat tagaagaatatcca	3027
Db	1315	-----GTGATCAAGAAATAATACCG	1335
QY	3028	actgatgatcttgaat tgtcgtcgtcgtgaagat tggggaagacaacacataa tgaatgtgc	3087
Db	1336	ACCGACGGGTCTGAGATCGTGTACTGTACCGAGGGGAGAGAACACGATTAAGCCCGTC	1395
QY	3088	tttaacgaggtcacaacaaatgtgtgcacatcatccacaagctgcgtttggggccctctga	3147
Db	1396	TTTCCGAGAGTGAAGCAAAACGAGCCATATATCCACACGCTGCGCTTGCGCCCTTCGGCA	1455
QY	3148	gtccaagaactagagagcctgtccaaaatgcagagaggttttaacagacatatgtctcaag	3207
Db	1456	GCGAAGGAGCTGAGAGAGCTGTGACAGATGACAGTGGATTTGGAGACATATGCTTCAAC	1515
QY	3208	caagttcagaacaat tgcctccat tgaatgtctttggggcccttcacagaa tggagct	3267
Db	1516	CAGCGCCGGAACAACGGCCATTTGAGCGCTTTTGGGGCCCTTTTCGTCGGGAACAGAGCT	1579
QY	3268	gtctctcagagcctcaccagccttgagagtgaaagagtttaacctccgaagaacagcagtg	3327
Db	1576	GCCTTCACGCGCTTCATTCACGCTGTGAAGTCAAGGCTTAAACCTTCAGACAAACAATGG	1639
QY	3328	atgaatggcacaagatcgttgcgaacgacacggttggaagaagacacttgttttctaac	3387
Db	1636	ATGAATGGCAAGTGGTGGTGAAGACGACGTGGGAAGACACTTTGTTTCTATCAAC	1695
QY	3388	tgaacatgtcag-----cctcccaaatcctctct--ggatcccaatgtgacagaagacaa	3441
Db	1696	TTGGAACGGAAGTTCCTTTCCCAATTCCTTTTGTGGGGTTCACGAGCGAAGACGACG	1755

Qy	3442	ggtgctcttgtagtgcgaacaaacccaataatggcctacctccaaatcccaagcatctg	3501
Db	1756	GACACTTTCCTTGTTGGAAAAACACAAAGATGGCTTACTTCCAAAGTCCCAAGCAAGGCG	1815
Qy	3502	aaggttgcacttggaaatacagctctgcgaagcaagctcacaaacttgaacctgactctg	3561
Db	1816	AAGGTGGCAGTGTGGAAATACAGCTTCAAAGCAAGTTCAACAAACCTCTGACTTGGACGTC	1875
Qy	3562	acgtcccgctgctgaatgcctacccctgcctcaatlaacagtagcttccaaacggaacag	3621
Db	1876	AGCTCCCGTGGTCCAGGCGCTACCTCTCCGTTACCGCTGACCTTAAATGAAACANA	1935
Qy	3622	gaaccagaacaaattccccaagcctctgtagttagtatagcgaataatctgcgaagagctcc	3681
Db	1936	GACACGGGCAAAATTCACAGCCCATGGTGTATTACACGAAGATTCACCAAGAAACCCG	1995
Qy	3682	ccaattctcaaggccagtgctacagccctgattgaaatcagtgaaatggaaaacagttacc	3741
Db	1996	CCCATTTTCAGGGCCAAAGTACAGCCCTATGTGAATCAAGAGATGGGAAACAGTTAC	2055
Qy	3742	ttggacactacgtgaataatgcagagtgctgcagctactcaagaatgaagttgcttactca	3801
Db	2056	TTGGAAATTACTGGAACAAAGAGCAGGGGCTGATGCTACCAAGATACGGTATCTACTCG	2115
Qy	3802	aggatttcacaactatgcgaagatggtgagatacagtgtaaaagtgcgggctctgcgga	3861
Db	2116	AGGTATTTTACAGCTTACGATGCAAAATGGTCGATTACAGCTTAATAAGTGGGCTCGGA	2175
Qy	3862	ggagttaacgcagccagacgagagatgataccccaagagatggagcactgttaactact	3921
Db	2176	GGAGTGAACACGCCACGACGAGAGGGCACCCCTCTGAGTGGACCTATGTACATACGT	2235
Qy	3922	ggctgatttggaaatgataatcaaatggaaatcccaagacctaaatlaataagat	3981
Db	2236	GGCTGGATTGAGAAATGGGGGAATCAAAATGAAACCCCGCAGACCCGACATTTAACAAAGAC	2295
Qy	3982	gatgttcaacacaagcaagtgtgtttcagcagaacatcctcggaggctcatttgbgct	4041
Db	2296	GACCTTTCAGGGCAAGCAAGTGTGTTCAGCGCAAGCCTCGGGGGGTTCAATTCGTGGCC	2355
Qy	4042	ctgtgctgtcccaaatgtcccaatacctgcattctcttcccaactctgcgaataacggaactg	4101
Db	2356	TCCGACGTTCCGAAGTCTCCCATACCCGATCTCTTTCCACCTCGTAAAGATCACGACCTG	2415
Qy	4102	aaggcggaatctcacgggggcagctcatatlaactgaacttgcagtcgaagctctcgggaatg	4161
Db	2416	AAGCGCGGATTCCAAAGGGGCAACCTCATACCTGACGTCGAGCGGCTCGGGGATGAT	2475
Qy	4162	tatgtaccatlgaaacagctccaagaatatataatcagataatgaatctcttgatctc	4221
Db	2476	TTCGACCAACGGAAGGCTGACAGGTATCATATGAAATTAAGCAAAATATTCTTGATCTC	2535
Qy	4222	agaagcaagcttcaatgaaatccttccaagtgaatlaactgcctctcatcccaagaagacc	4281
Db	2536	AGAGCAAGTTCANTGATTTCTGTTCAAGTGAACACACGATTTCTATCCAAAGAGGCGC	2595
Qy	4282	aactctgsgaagtccttgtgtttaaaccgaagaacattacttggaaaatgycacagat	4341
Db	2596	AACCTCAGAGGAGGCTTGTGTTTAAACCGAAGAGCATCCCTTTTCAAAATGGACAGAT	2655
Qy	4342	ctttcatctgtatctcagcgctgtgtgataaagtcgatctgaaatccgaataatccaaact	4401
Db	2656	CTTTTCATTTGCTGTACAGGCTGTGATTAACCAACACTGAAAGTCAAAATTCACAACAT	2715
Qy	4402	gaacagatctctgttattatctcccaagaccgcgcagagaacact---agtcctgat	4458
Db	2716	GCACAAAGTATCTTTGTTCCTTCCCGCGAGGCTCTTCGGAAGATCTCTCCGAAGACACT	2775
Qy	4458	gaacagctgtcctctgtcctataatataatcacaacagaacaaactctcgtgaaatcaat	4518
Db	2776	GCTCCTCTCTGCCTGTGCTGAAATTCAGGTCAACAGACACATTCCTGGAATTCACAT	2835

QY 4519 ttaaaattatgtgagatagaggaactgagctgtcaatagcctag 4569  
 |||||  
 DB 2836 TTTAAATATATGTGAGAGTGGAGAAATACGCTATCGATAGCCTAG 2886

RESULT 6  
 AB017156  
 LOCUS AB017156 2937 bp mRNA ROD 10-NOV-1999  
 DEFINITION Mus musculus gob-5 mRNA, complete cds.  
 ACCESSION AB017156  
 VERSION AB017156.1 GI:3721911  
 KEYWORDS GOB-5.  
 SOURCE Mus musculus adult intestine goblet cell CDNA to mRNA.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (sites)  
 Komiya,T., Tanigawa,Y. and Hirohashi,S.  
 Cloning and identification of the gene gob-5, which is expressed in  
 intestinal goblet cells in mice  
 JOURNAL Biochem. Biophys. Res. Commun. 255 (2), 347-351 (1999)  
 MEDLINE 99160866  
 REFERENCE 2 (bases 1 to 2937)  
 AUTHORS Komiya,T., Tanigawa,Y. and Hirohashi,S.  
 JOURNAL Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases. Tohru  
 Komiya, ERATO, JST, Genome Asymmetry Group, Doi Biosymmetry  
 Project, Tokodai 5-9-9, Tsukuba, Ibaragi 300-2635, Japan  
 (E-mail:tkom@bio.erate.trc-net.co.jp, Tel:81-298-48-1515, Fax:81  
 298-47-8901)

FEATURES  
 source  
 1. .2937  
 Location/Qualifiers  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /cell\_type="goblet cell"  
 /dev\_stage="adult"  
 /tissue\_type="intestine"  
 15. .2756  
 /gene="gob-5"  
 15. .2756  
 /gene="gob-5"  
 /protein\_id="BA33743.1"  
 /db\_xref="GI:3721912"  
 /translation="MESLKSPEFLILHLEGLVLSLSLIQLNNNGEGIVIALDHDP  
 EDEALIOHKIDWTOAPYLPFAATGRFYEKFNVAIIIPSMKAPETYPKLETFENA  
 DVLVSTSPGNDPEYTHIGAGEKGIHILPDJACKLKGVOPOORTFVHEVAH  
 FRMGVENEYNNDEKEFLSKGKPOAVKCSAATGKNVRRCCGSGCTTCKVCYDRVTC  
 LYKNCVFPYPPDHONKASIMRNQINSYVERTEKENHNOEAPNDONORCNLRSTVEY  
 IOESDEKQTPPTAOPAPPTSLIOIGRIVCLVDKSGSLINDRLNRMMNASRLF  
 LLOTVKOGSWAVTFDSAAVYVOSLQNSGADRLIKHLPTVSAGGTSICSGLRT  
 AFTVIRKKYPTDGSSEIVLLTDEEDNTISCFDIAVKSQGAIIHTVALGPAARAELEDS  
 KMTGLOIYSSDOVONNGLVDAFAALSGNAALIAQHSIOLESQVNLONNOMNGSVI  
 VQSSGKDFLFIITHTHTPTFTFIMDPSEVQNGFLLDITTKYAYLOVQGTAKVGRK  
 YSIOASQRTILFVTSRASATLPTTTPYVYNKNTGKPPSYTYAASIRQASPLR  
 ASYVALTESVNGKVTYVLELDWAGADATKNGSVSRPTAEPANRSVVKIIMALGV  
 TSDROAPAKNRAMIDSMIDGVRNMPPREPYSYVDKQLCFSTSSGSEFVATN  
 VPAAPIPDLPPPCQITDILKASIOGONLVNLTWTAPAGDIDGHRASNYIIRKSTSID  
 LRDHENTSLQVNTGLIPKESASEEIEFELGNTGPGNDIFIAIOAVDKSLKSEI  
 SNIRSVYIIPNOEPIPEDSTPCPDISINSTIPIQIHVILKIMMKLGEQVTLGH"

BASE COUNT 860 a 718 c 693 g 666 t

ORIGIN

Query Match 33.2%; Score 1517.6; DB 10; Length 2937;  
 Best Local Similarity 75.0%; Pred. No. 0;  
 Matches 2017; Conservative 0; Mismatches 554; Indels 117; Gaps 5;

QY 1888 aaggaatggtgaccagacatctgtatctgtgaagctacaggaagcgatttat 1947  
 |||||  
 DB 174 AAGGACATGCTGACTCAGCCCTCTCCATACCTTTTGAACCTACAGAAAGATTTTAC 233

QY 1948 ttcaaaaatgttgcatttctcctgaacaatcgtgaagaagaagcgtgactatgtgaga 2007

DB 234 TTTCAAAATGTTGGCATTGTTGATTTCCGACAGCTGGAGAGCCGGAAGCCTGATATACAGG 293  
 |||||  
 QY 2008 ccaaaacttgagacctcaaaaatgctgcatgtctgtgctgagctactccctcaagt 2067  
 |||||  
 DB 294 CCAAACTTGAAACCTTCMAAAAGCGTGTATGCTCTTTATACACACACACCTCTAGGC 353  
 |||||  
 QY 2068 aatgataaaccttaactgacgagatgaggaactgttgagagaagaaggatgaagatccac 2127  
 |||||  
 DB 354 AATGATGAGCCCTACACCGAATATATAGAGCATGTGTGAGAAAAAGGGATCAGATTAC 413  
 |||||  
 QY 2128 ctcaactcctatttcatcttgcaagaanaaagtgtcgtatataatgaccacaagtgaggca 2187  
 |||||  
 DB 414 CTGACTCTCTACTCTTTAGAGAGAAAGACCTGCTCAGTATGGGCCACAGACAGAC 473  
 |||||  
 QY 2188 ttgtccatgagtgagctcatcagatgaggagatcttgacagatcaataatgagag 2247  
 |||||  
 DB 474 TTTGTCATGATGCTGCTCATCTTCGATGGGAGCTGTTTATGAAATTCACAAACGACGAG 533  
 |||||  
 QY 2248 aaattcacttaccatggaagaatacaagcaglaagatgttcagaglatactgtgt 2307  
 |||||  
 DB 534 AAGTTCTACTTATCCAAAGCAAAACCCCAAGCAGTGTGAGGCTTACAGCATTACCGGT 593  
 |||||  
 QY 2308 acaaatgtatgaagaagtgtcagggagcagctgtaccacca---aagatgcacattc 2364  
 |||||  
 DB 594 AAAATATCAAGTTCCTCGTCCGTCACAGGAGGCAAGTTGTACATACGGAAGGTGTATATC 653  
 |||||  
 QY 2365 aataaataacagagactatgaaaagaatgtgagttgttcccaatcccgcaagag 2424  
 |||||  
 DB 654 GACAGAGTAAAGGAGGAGCTGTATTAAGCAATTTGTATTTGTACACAGATCCACACCAAAAC 713  
 |||||  
 QY 2425 gagaaggtctctataatgttttgcacaacatgttgatctatagttgaattctgtacaga 2484  
 |||||  
 DB 714 GAGAGGCTTCCATCATCATGTTTAAACCAAAATATCAATCTGTGTGTTGAATTCGTACAGAA 773  
 |||||  
 QY 2485 caaaacacacaagaagaagctccaacaagaacaaatataaaatgtcaattccggaagaca 2544  
 |||||  
 DB 774 AAAATATCAATCAAGCAAGAGCCCAAAAGCAAAACCAACGATGCAATTCGCAAGACAG 833  
 |||||  
 QY 2545 tgggaagtgtatccgtgtatcttgaggactttaagaaacacactctatgaacaacagcca 2604  
 |||||  
 DB 834 TGGGAATCATTCAGAGATGTGAGAGCTTCAACCAAAACCACTCCATGACAGCCACCA 893  
 |||||  
 QY 2605 ccaatccacactctcatatgtctgcagatgtgacaagaagtgtgtgtttagtctgtgac 2664  
 |||||  
 DB 894 CCTGACACCCACCTTCTCATCTGCTCAAAATGTGACAAAGAAATGTGTGCTTATGCTTGTAT 953  
 |||||  
 QY 2665 aaatctggaagcatgtgcgactgtgtaacgcctcaatctgactgataagaagccagctt 2724  
 |||||  
 DB 954 AAGTCCGGAGCATGTGAAAGCATGATCGCTTAACCGAATGATATGAGCAACGCGGCTT 1013  
 |||||  
 QY 2725 ttccctcgtgagaagaatgtgagctgaggggtccctgggttgaggatgtgacatgtgac 2784  
 |||||  
 DB 1014 TTCTCGTGCAGACTGTGAGAGGAGGATCTCGGCTGGGATGTGTGACTTTGACAGTGT 1073  
 |||||  
 QY 2785 gcccatgtacaaaatgaactacatacagataaacaagltgagcagtgagaggaacacactggcc 2844  
 |||||  
 DB 1074 GCTATGATCAAAAGCAAGCAAAAGATTAAACAGTGTGCTGACAGAGATCTGCTGATC 1133  
 |||||  
 QY 2845 aaaaagattactcgtcagcagacttcaaggaaggaagctcactctgcagcgaggtctgacatggca 2904  
 |||||  
 DB 1134 AAGCATTACCCACAGTATCTGACAGGAGGACATCTATATGCTTGTGCGCTTTCGAGACGA 1193  
 |||||  
 QY 2905 ttactgatatgtgycacaactttgcctgtttcatalgacacacagcagttatggygagtg 2964  
 |||||  
 DB 1194 TTTAC----- 1198  
 |||||  
 QY 2965 cgaacaagaatccaaatltggcctctctgctgagctgtgactgtatgtagtgaagaataat 3024  
 |||||  
 DB 1199 -----ATTGATTAAGAAGAGTAT 1217  
 |||||  
 QY 3025 ccaactgtatgagatcgtaaatgtgctgctgagcagatgtggggaagaacaacactaaagtgg 3084  
 |||||

Db 1218 CCAACTGATGATGTAATTTGCTGCTGACCGATGGGGAGACACACATTAGACAGC 1277  
 QY 3085 tgccttaacagaggtcaacaagaatggtgcatcatccacagatgcgtttggggccct 3144  
 Db 1278 TGCCTTGAACCTGGTGAACAGACAGGGGCCATCATCCATACAGTGGCCCTGGAGCGGCT 1337  
 QY 3145 gcaactcaagaactagaggaagctgtccaaaaatgacaggaagtttacaacatatgtctca 3204  
 Db 1338 GCCCCTAAAGAGCTTGAACAGACCTGTCCAAATGACAGGAGCGCTGCACACATACCTTTCG 1397  
 QY 3205 gataaagttcgaacaatagccctcatgtatgtcttttgggccccttcatcagaatbga 3264  
 Db 1398 GATGAGGTTCAACCAATGAGCTTGTGATGCTTCCGACACTCTCTCGGAATATCGC 1457  
 QY 3265 gctctctcctcaagctccatccagcttgaagtagaaggaattaaacctccaaagaacag 3324  
 Db 1458 GCGATCGCTACGACCTCCATCCAGCTGGAGACAGGGAGATTATCTCCAGAAATACCAA 1517  
 QY 3325 tgaatgaatgacagatgatalcgtgacagacgctgggaaaggaacttcttctatc 3384  
 Db 1518 TGAATGATGGCTCAGTATGATGAGACAGCTGGTGGGCAAGACACTTGTCTTATTC 1577  
 QY 3385 accctggacaatgacagctctcccaaatcctctcgtggatcccaagtgcagaagaagct 3444  
 Db 1578 ACCCTGGACCAAGCATCCCTCTACAAATATTTATCTGGGATCCAGGGAGTGAACAAAT 1637  
 QY 3445 gacctttagtgggcaaaaaaaccaaaatggcctacccctcaaatcccaagatgtctag 3504  
 Db 1638 GGTTTTATCTAGACACACACACTAAGTGGCTTACCTCCAACTCCAGGACGGCTAAG 1697  
 QY 3505 gttggcacttggaaatcacagctctgcaagaactcaacaaccttgaaccttacttgcag 3564  
 Db 1698 GTGGCTTTTGGAAATACAGATTCACAGCGAGCTCACAGACTCCACTTGCATGTCAC 1757  
 QY 3565 tcccgctgctcaatgtcatccctgcctccaaattacagtgtactccaagaacagaaagac 3624  
 Db 1758 TCCCGTGACGAAGTGTCTACACTGCTCTCTATTAAGTACCGCCGTTGTAATTAAGAC 1817  
 QY 3625 accgcaaatcccccagccctctgtagtattatgaatattcggcaagaagccctccca 3684  
 Db 1818 ACAGGGAATTTCCCGAGCCCTGTACAGTGTATGACAGCATTTCCGCAAGAGACCTCGCT 1877  
 QY 3685 attctcagggcagtgltcaacagccctgtatltgaatcgtgaatggaaagaaacttactt 3744  
 Db 1878 ATTCTCAGGGCCAGCGTACAGCCTTGATGATCTGTGANTGGAATAAACGATACCTG 1937  
 QY 3745 gaactactgataatgtagcagagtgctgtatgtactaagaatgtagcgtgtctactcaag 3804  
 Db 1938 GAATTACTGGATTAACGAGCGAGTCCCATCCACCAAGAAATGATGGTCTACTCAAG 1997  
 QY 3805 tatttcaactttagacagaatgtgtatgatacagtgtaaaagtggcgctctgggaaga 3864  
 Db 1998 TTTTTCACAGCTTTTGAATGCAATATGATGATACAGCTTAAATATGAGGCTCTGGAGGA 2057  
 QY 3865 gttcaacgacagcaggaagatgatalacccagcagagtggaagacatgtacataactgttc 3924  
 Db 2058 GTCACTTCAGACAGACAGAGAGCAGCACTCCGAAAGAACAGCCATGTACATATAGGC 2117  
 QY 3925 tggatggaatgtatgataatcaatgatacccaagaactgaaatgaagaatgac 3984  
 Db 2118 TGGATTTGAGGTGGTGAATGAATGAACCCACACCTCTCTGAAACATGTT-----AT 2171  
 QY 3985 gtcaacaacagcaagatgtgttctcagagaacatcctcggagaggtctgtgtgctct 4044  
 Db 2172 GTTCAAGACCAAGACAGCTTGTCTTCAGAGACATTTTCAAGGGGATTCGTTTGTGGCCAC 2231  
 QY 4045 gatgtcc---caatgtcccatatactatctctcccaactgycgaatcaaccagact 4101  
 Db 2232 AATGTCCCGAGAGAGAGCTCCATTCCTGACCGCTTTCACACCTGTCAATATCACTGACCTG 2291  
 QY 4102 aagcgagaatctcagggggcagctcatatcactgtactgtgacagctcctcgtggagatgac 4161  
 Db 2292 AAGGCACACATCCAAAGGCGAGAACCTGGTGAATCTGACGTGACGGCTCTGGGGATGAC 2351

QY 4162 tatgaccatggaacagctcacagaatatactatcgaatgaagaatcttgcac 4221  
 Db 2352 TACGACCAAGGAGAGAGCTTCCACTACTCATTCGAATGACAGACAGATGTGTTATCTC 2411  
 QY 4222 agagaacattcaatgaaatctcttcaagtgaatactatgctctcatalcccaaggaagcc 4281  
 Db 2412 AGGACACTTCAACACCTCACTCACTCAAGTGAACACTACCGGTATATCCCAAGAGGCC 2471  
 QY 4282 aactctggaagatcttcttctttaaaccagaataactacttcttgaagaatggcagat 4341  
 Db 2472 AGCTTGAGGAATCTTGTGAGTTGAAGTTGAAGTGAAGCAACACTTTTGAATAAGGACAGAT 2531  
 QY 4342 ctcttctgctatcgaagcgtgtatgaagtgatgcgtgaatacgaataatccaaact 4401  
 Db 2532 ATCTTCATTCCTATTCAGAGCTGTGATGATGATGATGATGATGATGATGATGATGAT 2591  
 QY 4402 gcaagagatcttltgttattctcctccacagactccgcagagacactagctcgtatgaa 4461  
 Db 2592 GCACGGGTGCTGTGTTCATCCCGCTCAG-----GAGCCGCCATTCGCCAAGAC 2642  
 QY 4462 acgtctgctcttctccaaatataatcaacagacacactcctgcacatcattta 4521  
 Db 2643 TCAACTCCCGCTTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2702  
 QY 4522 aaaatattgtaagtgataggaagactgcagctgtcaatagcctag 4569  
 Db 2703 AAGATATGTGGAAGTGGCTAGGGAATGCAAGTGAACATGATGATGATGATGATGATGAT 2750

RESULT 7  
 AF127035  
 LOCUS  
 DEFINITION Homo sapiens calcium-activated chloride channel protein 2 (CaCC2)  
 ACCESSION AF127035  
 VERSION AF127035.1 GI:5726288  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 3204)  
 AUTHORS Agnel,M., Verma,T. and Culouscou,J.M.  
 TITLE Identification of three novel members of the calcium-dependent chloride channel (CaCC) family predominantly expressed in the digestive tract and trachea  
 JOURNAL FEBS Lett. 455 (3), 295-301 (1999)  
 MEDLINE 99364503  
 PUBMED 10437792  
 REFERENCE  
 2 (bases 1 to 3204)  
 AUTHORS Agnel,M. and Culouscou,J.-M.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-FEB-1999) Groupe Genomique, SYNTHELABO, 10 Rue des Carriers, Rueil-Malmaison 92500, France  
 FEATURES  
 source  
 1..3204  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="colon"  
 1..3204  
 /gene="CaCC2"  
 29..2782  
 /gene="CaCC2"  
 /note="bovine epithelial chloride channel homolog"  
 /codon\_start=1  
 /product="calcium-activated chloride channel protein 2"  
 /protein\_id="AAD48398.1"  
 /db\_xref="GI:5726289"  
 /translation="MGLEFRGVFLVLCILHQSNTSFKLNKNGFEDIVIVYDPSVE  
 DEKTIIEDVNTASTYLEFATKREFPFKNSVSLIDENKVENQYVRPHENKRD  
 VIVAPPLGDEPDTKQPECEGKEVIEHTPDLLGKQNEVGPGLVHMAHL  
 RMGVFDEYNEQDPYRAKSKIEATRCASIGISGRNRYVKCGGSGLSLRACRIDSTFKL

YKRCQCFPPHVOUEKASIMFQMSIDSVYEFQCNETHNOEAPSLONIKCNRSTWBI  
 NSBDEENITPWPPEPPYFSLKTRQIVCLVLDKRSQSMGKORLRMMQAKHFL  
 LQYENESWQWGHPESTAIYVKKLQJISQBSERTLWAGTEPPLGTSJSCISQKIA  
 FOYVIGELYSQDSEVJLIDGDLGNASSCIDEVQOSAIYHFLINAGAEADYEMSI  
 KITGHSFYHQSDEKONGGLIDAGDLNCSLQSOLOSKEGTLTANMAMNDVYI  
 IDSVYKCDTFFPLTWNLSPLSILMDPSTIMBNPTVDATSKMAYLSIPGAKGTWIA  
 YNLOAKANPELITIVYISVRANSSVPTIVYAKMKNDVSPYENYIAELOSQYV  
 LGANVYPTIESQNHPEVLELWNGAGSPNDVSKYFPEYENYIAELOSQYV  
 GANVYATRLKRLPINRAYIPGVYVNEITFANPREIDEDTOITLDESPFASGAV  
 ISQYVSPDLQPPSOIYTDVLEADASKFMEKENTISEENATHIPLAKSIDKSNLTK  
 DLROSEFDALQVNTDLSPEKASKSFMEKENTISEENATHIPLAKSIDKSNLTK  
 VSNIAQVTLFIPANQEDIDPPTPTPPTRDKSHNSGVNISTVLVTSQVYVNFILS

Db	1953	TTTGTGAACCTTTGGATATATGGTGCAGGCGCTGATTTCTTTCAGAAATGATGAGCTACT	2012
QY	3800	caaggatattcaacaactatgacacgaatggtlagaatacagltgaanaagtcgggctcttg	3859
Db	2013	CCAGGTATTTTACAGCAATATACAGAAATGCGACATATAGCTTAAAGTTGGGGCTCATG	2072
QY	3860	gaggagtttaagcgacgacgaagggagtgatatacccgaggaugtgagcactgtaatatc	3919
Db	2073	GAGGAGCAAAACACTGCCAGGCTTAAATTAACGGCTCTCCACTGTAATAGACGGCGCTACTAC	2132
QY	3920	ctcgcgtgattcgaagatgatgaataacaaatgaaatccacaagaacctgaaataataag	3979
Db	2133	CAGGTGGTGTAGTGAAACGGGAAATTTGAAGCAAMCCGCCAAGACGTGAATTTGAT---G	2189
QY	3980	atgatggttcaacacaaagcaagtggtgttccagcagaacatacctcggagctcatattgtg	4039
Db	2190	AGGATACACACACACACCTTTGGAGAGATTTCACCCCAACAGCATCCGGAGGTGCAATTTGTGG	2249
QY	4040	cttcgcgtgcccgaatgctcccaatacctgatactcctcccaacctggcacaataccagc	4099
Db	2250	TATCACAAGTCCCAAGCTTCTCCCTTGCTGCTACGCATATCCACCAAGTCAATATCAGACC	2309
QY	4100	tgaagcggaataatcacaaggggcagctcgaatctcaatctcgaacttgacacgctcctgggagt	4159
Db	2310	TTTGATGCCACAGTTCATGAG---ATAAGATTATTTTACATGATGACGACGACGAGGATA	2266
QY	4160	attatgacctggaagcagctcacaagaatatacatctgaataagaagacaaagtatcttgatc	4219
Db	2367	ATTGTGATGTTGGAAAAATTCACAGTTATATCAATAAATAAGTGAAGTATTCCTTGATC	2426
QY	4220	tcaagacaagttcaatgaaatctcctcgaagaaatactgctcctcatccacaagaag	4279
Db	2427	TAAAGACAGTTTGATGATGATGCTCTTCAAGTAATTAAGTACTGATCTGTCCACCAAGAGAG	2486
QY	4280	ccaactcgtgaggaagctcttttgttttaaacacagaacatactctttgaaatgacag	4339
Db	2487	CCAAATCCCAAGAAAGCTTTTGCAATTTTAAACAGAAATATCTCGAAGAAATGACACC	2546
QY	4340	atcttcattatgcatcacaagcgtgtgataagcgcgtctgaaatcgaataatccaaca	4399
Db	2547	ACATTTTATTTAGCCATTAAAGTATAGTATTAAGCAATTTGACATCAAAAGTATCCANCA	2606
QY	4400	ctgcacgagatccttgtttatcttcctccacagactccgcagagacacctaagctcgtatg	4459
Db	2607	TTGCACAAGTAACTTTGTTATTCCTCAAGCAAAATCTGATGACATGATCACTCACTCA	2666
QY	4460	aaagctgcgtccctgtgcctaattcata	4489
Db	2667	CTCTCACTCTCACTCCCTGATTAAGTCA	2696
RESULT	8		
AK000072			
LOCUS	AK000072		
DEFINITION	AK000072	3221 bp. mRNA	22-FEB-2000
ACCESSION	AK000072		
VERSION	AK000072.1		
KEYWORDS	oligo capping; fis (full insert sequence);		
SOURCE	Homo sapiens colon cDNA to mRNA, clone_1lb:COL clone:COL01613.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y.,		
	Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,		
	Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.		
TITLE	NED0 human cDNA sequencing project		
JOURNAL	Unpublished (2000)		
REFERENCES	2 (bases 1 to 3221)		
AUTHORS	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,		
	Shibahara,T., Tanaka,T. and Nakamura,Y.		

FEATURES	COMMENT
JOURNAL	Direct Submission
	Submitted (13-FEB-2000) to the DDBJ/EMBL/GenBank databases. Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)
	NEBO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).
SOURCE	Location/Qualifiers
	1. 3221
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="COL01613"
	/clone_1lb="COL"
	/tissue_type="Colon"
	/note="Cloning vector pME18SFL3"
	1. 3221
	/note="Highly similar to ECCL_BOVIN EPITHELIAL CHLORIDE CHANNEL PROTEIN"
BASE COUNT	1105 a 600 c 634 g 882 t
ORIGIN	

Query Match	24.7%	Score 1129.8	DB 9	Length 3221
Best Local Similarity	67.1%	Pred. No. 1.1e-279		
Matches 1749; Conservative	0	Mismatches 747	Indels 111	Gaps 6

Oy	1889	agagcagcagcagccagcagcctcctcctatctcttgaaagcctcagaagagattctat	1948
Db	200	AGGATATGGTGACTACAGCTTCTACGTACTGTTTGAAACCACAAAAAAGATTTT	259
Oy	1949	tcaaaaatgttgcattcttgatcttcctgaaacatgtaagaacaaagctgactatgagac	2008
Db	260	TCAAAAATGTAATTAATTAATTCCTGAGAAATTGGAAGAAAAATCCTCAGTACAAAAGC	319
Oy	2009	caaaacttgagacctacaanaatgctgtagtcttcggtctgcgtagtctcctccagta	2068
Db	320	CAAAACATGAAACCAATMAACATGCTGAGATTATTAATGTCACACACTACCTCCAGGTA	379
Oy	2069	atgtagaaccttacaagcagatgtagcgaactgtgagagagaagggtbaaagatccacc	2128
Db	380	GAGATGAGACCATACCAACAGCACTTACAGAAATGTGGAAGAAAGCAATACATTCACT	439
Oy	2129	tcaactctgatttcatctgacagaaaaaagtctgctgaatatgacacaaagctaggagat	2188
Db	440	TCACCCCTGACCTTCTACCTTGGAAAAAACMAATGAATATATGACCACCAAGCAACGT	499
Oy	2189	ttgtccatgagttgggctcactcctcagatggggagatttgaagagtacaaatgatgaga	2248
Db	500	TTGTCCATGAGTGGGCTCACCTCCGGTGGGAGGTGTTTGATCAGTACATGAAGTCAAGC	559
Oy	2249	aattcctactatccaaagaagaatc-----aaagcgtaaagatggttcagcagatctactg	2305
Db	560	CTTTCTCTCCGTGCTAACTCAAAAAAATGCAAGCAACAGAGTGTTGCCAGAGTATCTTG	619
Oy	2306	gtacaaatgtgagtaagaagtgtcagggagcagcgtgttaccccaaaagatgacatca	2365
Db	620	GTAACAATFAGAGTTTAAATAGTGTCAAGGAGGCGCTGCTTAGTAGACATGCAAGAAATG	679
Oy	2366	ataaagtacaagacctatgaaaaaagatgtgagttgtttctccaatcccgccagagc	2425
Db	680	ATTCTACACAAACACTGTATGGAAGAAAGATTGTCAATTTCTTCCTGATAACTACAAACAG	739
Oy	2426	agaagcgtctataatggttgccacaacatgctgattctatgatttgaattctgtcacaag	2485
Db	740	AAAAAGCATCCATTAATGTTTATATGCACAAATATGATTTCTGTTCTTAATTTTGTAAACGAA	799
Oy	2486	aaaaaccacaagaagctctccaaacaagcaaatcaaaatgcaatctccgaagcaat	2545



Db 800 AAACCCATTAATCAGAGCTCCAGCCCTACAAAACATTAAGTCGAATTTTGAAGTACAT 859  
Qy 2546 gggagatgacccgtgacatctgagactttaagaanaaccactccatgacacagacac 2605  
Db 860 GGGAGGATGATTAAGCAATTTGAGGATTTTAAAAACACCATACCATGCTGACACACCTC 919  
Qy 2606 caaatccaccctctcatctgctgcagatctgacaagaatctgctgctgctgaca 2665  
Db 920 CTCACCTGCTTCTCATCTGCTGAGATCCGTCGAAGAATGTGTGCTAGTTCTTGATA 979  
Qy 2666 aactcggaaacatgagcagcgtgtaaccgctcactgacatgacatgacacagcagctt 2725  
Db 980 AGCTGGAACCATGGGGGTAAGGACCGCTTAATCGAATGAATCAAGCAGCAAAACATTT 1039  
Qy 2726 tccctgcgcagacagctgagcgtggtggtggtggtggtggtggtggtggtggtggt 2785  
Db 1040 TCCTGCTGCAGACTGTTGAAATGATCCTGGGGGATGCTTACATTTGATGATGACTG 1099  
Qy 2786 cccatgtacaaaatgaaatcatatagataaaacagctgagcagctgacagagacacactgc 2845  
Db 1100 CCACCTATTGTAATTAAGCTTAATCCAAATTAAGCAAGCTGATGAAGCAACACTCATGG 1159  
Qy 2846 aaagattactcgcagcagctcgcagcagcagcagcagcagcagcagcagcagcagcag 2905  
Db 1160 CAGGATTACTACATATCTCTGGAGGAACTTCATCTGCTGCAATTAATATATCAT 1219  
Qy 2906 ttactgatatgtggaacactgtgctgttcttccatgacacacagcagcttctggtggtgc 2965  
Db 1220 TTTCAGTGATTTGGAGAC----- 1237  
Qy 2966 gacaaagaaatccaaatctggcctctctgctgcagcagctgagatgagagaataatc 3025  
Db 1238 -----TACATTCCTC 1246  
Qy 3026 caactcgtatctgaaatgtgctgctgacgagatgggagaaacacataatgagct 3085  
Db 1247 AACCTCGATGATCCGAGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306  
Qy 3086 gcttcaacgagctcaacaaatgctgctcactccacacagctgctggtggtggtggtggt 3145  
Db 1307 GTATGTGATGAAGTGAACAAAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1366  
Qy 3146 cagctcaagaactagagagcgtccaaatgacagagagcttaccagacatgctcag 3205  
Db 1367 CTGATGATGAAGTGAATGAGATGAGCAAGTAAACAGAGGAGCAATTTTATGTTCTGAG 1426  
Qy 3206 atcaagcttcaagaacatgctcactgactgcttctggtggtggtggtggtggtggtggt 3265  
Db 1427 ATGAAGCTCAGAAACAAAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1486  
Qy 3266 ctgctcctcagcgtccatccagctgagagtaagatlaaacctccagaaacagcagct 3325  
Db 1487 ATGCTGCCAGAGAGTCCCTTCAGCTGCAAGTAAAGGATTAACACGTAATGATATGCTCT 1546  
Qy 3326 ggaatgaatgacagctgctgctgacagcagctggtggtggtggtggtggtggtggt 3385  
Db 1547 GGATGTAACGACACTGCTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 1606  
Qy 3386 cctggagcaatgacagcgtcccaaatctctctctggtggtggtggtggtggtggtggt 3445  
Db 1607 CATGGAACATGCTGCCCTCCAGTATTTCTGTGGGATCCACAGTGAACATTAATGAAA 1666  
Qy 3446 gcttctgagtgagcaaaaacacaaaatgctcactccaaatcccccagcttgcag 3505  
Db 1667 ATTTCACAGTGAATGCAACTTCCAAAATGCTATCTCATATTTCCAGGAACATGCAAAAG 1726  
Qy 3506 ttggcactctggaaatcagctctgc-----aagcaagctcacaacaccttgcacctgactg 3559  
Db 1727 TGGGCACTTGCGCATACATCTTCAAGCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1786  
Qy 3560 tcaagctccgctgctcacaatgctcactccaaatlaacagctgactccaaaagac 3619

Db 1787 TAACCTCTGAGCAGCAAAATTTCTGTGTGCTCCCATCATCAGATGCTAAATGAATA 1846  
Qy 3620 agacacacagcaaatctccaccgctctggtggtggtggtggtggtggtggtggtggtggt 3679  
Db 1847 AGGACGTAAACAGTTCCTCCAGCCCATGATTTGTTTACGAGAAATTTCTCAAGGATATG 1906  
Qy 3680 ccccaatctcagggccagctgctcagcagcctgactgaaatgaaatgaaatgaaatgaaat 3739  
Db 1907 TACCTGTTCTTGAGAGCCCAATGCTACTCTTTCATTTGAATGACAGAAATGACATACAG 1966  
Qy 3740 ccttggacactcgtgaaatgagcagcgtgctgactgactgactgactgactgactgactg 3799  
Db 1967 TTTTGGAACTTTTGGATAATGTCAGGCGCTGATTTCTTCAAGAAATGATGAGTCTACT 2026  
Qy 3800 caagatctcacaactatgacagatgtagatacaagtgtgaaagtgtggtggtggtggtggt 3859  
Db 2027 CCAAGTATTTTACAGCATATACGAAATGCGAGATATFACCTTAAAGTTGCGGCTCATG 2086  
Qy 3860 gaagagtltaacgacagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 3919  
Db 2087 GAGGAGCAAACTGCTCCAGCCTAAATTAACGCTCCACTGAAATGAGCGCGCTACATAC 2146  
Qy 3920 ctgctgacttggagatgagatgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaa 3979  
Db 2147 CAGGCTGGGTAGTGAACGGGAAATGTAACCAACCGCCCAAGACCTGAATTTGAT--G 2203  
Qy 3980 atgagtctcacaacagcaagctgcttctcagcagacacactcctggagagcctctgtg 4039  
Db 2204 AGGATACTGAGACCACTTGGAGGATTTTACGCCGAAACAGCATCCGAGATGCTATTTGTG 2263  
Qy 4040 ctctctgctgctcccaatgctccacatcctcctcctcctcctcctcctcctcctcctcct 4099  
Db 2264 TATCACAAGTCCCAAGGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2323  
Qy 4100 tgaagcggagaaatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 4159  
Db 2324 TTGATGCCACAGTTCATGAGG---ATTAAGTATTTCTTACATGACACGCCAGAGATA 2380  
Qy 4160 attatgaccaatgaaacagcctcacaagatatactcgaatgaatgaatgaatgaatgaat 4219  
Db 2381 ATTTTGTGTTGGAAAGTTCAACGTTATATCATTAAGAAATGATGCAAGTATTTCTGATC 2440  
Qy 4220 tcaagagcaagctcaatgaaatctctcctcctcctcctcctcctcctcctcctcctcct 4279  
Db 2441 TAAAGACACTTTTGAATGATGCTCTTCAAGTAAATCTGATGATCTGCTGCTGCTGCT 2500  
Qy 4280 ccaactctggagaaagcttcttcttcttcttcttcttcttcttcttcttcttcttctt 4339  
Db 2501 CCAACTCCAAAGAAAGCTTTCATTTAAACAGAAATATCTCAGAAAGAAATGCAACCC 2560  
Qy 4340 atcttctcatctgctatcagcgtctgctgataagctgactcgaatcagaatataccaca 4399  
Db 2561 ACATATTTATTTGCAATTAATAAGTATAGATTAAGCAATTTGACATTAATAAGTATCCA 2620  
Qy 4400 ttgcagagatctctgttcttcttctcctcctcctcctcctcctcctcctcctcctcct 4456  
Db 2621 TTGCACAGATTAATCTTGTATATCTTCAAGCAAAATCTGATGACATGATGATGATGAT 2680  
Qy 4457 atgaacgctgctgct 4483  
Db 2681 CTCCTACTCTACTCTCTACTCTCTGATA 2707

RESULT 9  
AX092338  
LOCUS AX092338 3265 bp DNA PAT 21-MAR-2001  
DEFINITION Sequence 69 from Patent W00116318.  
ACCESSION AX092338  
VERSION AX092338.1 GI:13444483  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 3265)  
AUTHORS Eaton,D.L., Filvaroff,E., Gerritsen,M.E., Goddard,A.,  
Godowski,P.J., Grimaldi,C.J., Gurney,A.L., Watanabe,C.K. and  
Wood,W.I.  
TITLE Secreted and transmembrane polypeptides and nucleic acids encoding  
the same  
JOURNAL Patent: WO 0116318-A 69 08-MAR-2001;  
Genentech, Inc. (US)  
FEATURES Location/Qualifiers  
source 1..3265  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 1159 a 596 c 632 g 878 t  
ORIGIN

Query Match 24.7%; Score 1129.8; DB 6; Length 3265;  
Best Local Similarity 67.1%; Pred. No. 1,1e-279;  
Matches 1749; Conservative 0; Mismatches 747; Indels 111; Gaps 6;

QY 1889 aggcagatggtgaccagcagcctctgtatctcgtcttgaaagctaacaggaagcatttatt 1948  
DB 182 AGGATATGCTACACACCTTCTACGTACCTGTTGAAGCCAGAAAAAGATTTTT 241  
QY 1949 tcaaaaatgttcacattctgtatccctgaacatgaaagcagcagctgactatgtgagac 2008  
DB 242 TCAAAATGTCTATATTAATTCCTGGAATTGGAAAGAAATCCTCAATGACAAAGC 301  
QY 2009 caaaactgagacctacaaaaatgtcgtatgtctgtcgtcgtacgtactcctccagta 2068  
DB 302 CAAATGTAAGAAACATTAACATGCTGATGTTATGTGACACACCTACCTCCAGGTA 361  
QY 2069 atgtgaacctaacatgagatgagcagcactgtgagagaaaggtgaagaagcacc 2128  
DB 362 GAGATGAACCTACACCAAGCAAGTTCACAGATGTGAGAGAAAGCGAATATACCTACT 421  
QY 2129 tcaatcctgattcatctcagaaaaaagttcagctgaatctgagcacaagtgagcatt 2188  
DB 422 TCACCCCGACCTTCTACTTGGAAAAAACAAATGATATGACACACGCAAACTGT 481  
QY 2189 ttgtccatgagtggtgctacatcagatgaggaagtaattgacgagtacaaataatgaga 2248  
DB 482 TTGTCCATGAGTGGGCTCACCCTCGGTGGGAGTGTGATGATGACAAATGACATCAGC 541  
QY 2249 aatctactatccaatggaagaatac---aagcagtaagatgttccagcaggtattact 2305  
DB 542 CTTCCTACCGTCTAAGTCAAAAAAATCGAAGCAAGGTTGCCAGGTATCTCTG 601  
QY 2306 gtacaatgtagtaagaagatgtcagggagcagctgttacaacaaagaatgacattca 2365  
DB 602 GTAGAAATAGAGTTTATATAGTGTCAAGAGGAGCTGTCTTAGTAGACATGCAAAATTG 661  
QY 2366 ataaagtaacagagactatgaaaaaagatgagatgttcttccatcccgccagcag 2425  
DB 662 ATTCTACACCAAACTGATGGAAGATGTGCAATCTTCCTGATTAAGTCAAAACAG 721  
QY 2426 agaagcgtctataatgtttgacacacatgttgcattatagttgaattcgttaagaac 2485  
DB 722 AAAAAGCTCATATATGTTATGCAAAAGTATGATTTGTTGTAATTTGTAAAGAA 781  
QY 2486 aaaaacacaacaagaactccaacaaagcaaatcaaaaatgcaatccgaaacacat 2545  
DB 782 AAACCCAAATACAGAACTCCAAACCTCAACAAACATTAAGTGCATTTTAAATATACAT 841  
QY 2546 gggagatgataccgtgattctgaggaacttaagaanaacacactctatgacaacagcac 2605  
DB 842 GGGAGGTATAGCAATCTGTGAGATTTTAAAAACACCATCCCTGCTGACACCCCTC 901  
QY 2606 caaatccacacttcatctgctgacatgtgacaagaagaatgtgtttagtcccttgaca 2665  
DB 902 CTCACCTGCTCTTCATCTGCTGGAAGATCAGTCAAGAAATGTGTGCTTAGTCTTGATA 961

QY 2666 aatctggaagcatggtcgactgtgtaaccgcctcaatccgactgaatcaacagcagcatt 2725  
DB 962 AGCTCGAAGATGGGGGCTTAAGACCGCCTTAATTCGAATATCAATCAACACCAAAACTT 1021  
QY 2726 tctgtctgacagcatgtgactgtggtgtccgtgtgttggaatgtgacatttgacgtgct 2785  
DB 1022 TCTGCTCCAGACCTGTTGAATGATGATCTGTTGGGATGTTCACTTATGATGACTG 1081  
QY 2786 cccatgtcaaaaatgaactctacagataaagcagtgagcagtggaagcagcactcgaca 2845  
DB 1082 CCATATTGTTAAATAGCTTAATCCAAATTAAGAGATGATGAAGAAACACACTCATAG 1141  
QY 2846 aagaattacccagcagccttcagagaggaagcctccatctgacgagcgttcgactcgat 2905  
DB 1142 CAGATTTACCTACATATCTCTGSGAGAGACTTCATCTGCTTGGAATTAATATGAT 1201  
QY 2906 ttaactgatatgtgcaaacatttgcctgtttccatgacacacagcatatagggagtg 2965  
DB 1202 TTCAGGTGATTTGAGAGC----- 1219  
QY 2966 gacaagaataacaaatgtggcctcctgcgtcagcttagttagtaagaataatc 3025  
DB 1220 -----TCAATTC 1228  
QY 3026 caactgtagatctgaaattgtctgtcgtcagcagatggggaagacacataaagtgg 3085  
DB 1229 AACTCGATGAGATCCGAATACTGCTGCTGACGTAGGGAGAGATTAACCTGCAATTCCT 1288  
QY 3086 gcttgaacaggtcaaaaaagttgttcacatccacacagtcgcttggggcccttg 3145  
DB 1289 GTATGATGAAGTGAACAAAGTGGGCGCATTTGATTTATTCCTTTGGGAAGAGCTG 1348  
QY 3146 cagctcaagaactgagagcgtgtccaaaatgacagggagtttaagacataatgtcag 3205  
DB 1349 CTGATGAAGCAGTAAATACAGATGAGACAGATACAGGAGAGTAAATTTATGTTAG 1408  
QY 3206 atcaagttcagaacaatggtccctcattgtctcttggggcccttcacacagaaatggag 3265  
DB 1409 ATGAGCTTCAGAACAAATGGCCATTTGATGCTTTGGGGCTTTACATCAGGAATACCTG 1468  
QY 3266 ctgtctcagcgtccatccagcttgagagtaaggatlaacccctccaagaacagcag 3325  
DB 1469 ATCTCTCCAGAGTCCCTTCAGCTCGAAGTAAGGATTAACCTGATTAATGCTT 1528  
QY 3326 ggaatgaagcaatgtagtccgtgagacacacogtggaaagacacttgttctatca 3385  
DB 1529 GGATGAAGGACCTGTCAATTAATGATGATGACGTGGAAAGGACAGCTTCTTCATCA 1588  
QY 3386 cctggaacaatgagcctcccaaatccctctcgttgatcccgatggagacagaagtg 3445  
DB 1589 CATGGAACAGTGTGCTCCAGATATTTCCTCTGTGGATCCCGATGGAACAAATAGGAA 1648  
QY 3446 gctttagtgtagacaanaaacacaaatgtgctcactccaaatcccgagcatctgaag 3505  
DB 1649 ATTTCAGATGATCAACTTCCAAAAATGGCCTATCTAGTATTCAGAGAACATGCAAGG 1708  
QY 3506 ttggcacttgaataaagctgcg-----aagcagctcaaaaacttggccgagc 3559  
DB 1709 TGGGCATTTGGCATATCAATCTTCAAGCCAAAGCCCAACCCAAACATTAATATTACG 1768  
QY 3560 tcaagctccgtgtcgtccaaatgtccacgtcccaatlaagatgactccaaaagaaca 3619  
DB 1769 TAACTTTCGAGACCAATATTTCTGTGCTCCCAATCAAGTGAATGTAATTAATGAATA 1828  
QY 3620 aggaacagcaaaattcccgagcctctgtgattatgtaaaatctgcgaaggagcct 3679  
DB 1829 AGGAGCTAAACAGTTTCCCGACCCCAATGATGTTTACGAGAAATTTCTACAGGATATG 1888  
QY 3680 ccccaactcgaagcaggtgtcagagcccgatgaaatgaatgaatgaatgaatgaatga 3739  
DB 1889 TACCTGTTCTTGAGACCAATGAGTGTCTTCTATGAAATCAAGATGACATTAAGAG 1948  
QY 3740 ccttgaactactgataatgagacaggtgtctgatactactaagaatgacggtgtact 3799

```

Db      1949 TTTTGGAACTTTTGGATTAAGTGTCAGGCCCTGATTTCTTTCAAGAAATGAGTACTTACT 2008
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      3800 caaggtatcacaactatgacacgagatgtagatcacgtgtgaaagtcgagccctcgtg 3859
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2009 CCAAGTATTTTACAGCATATACAGAAATGCGACATATACCTTAAAGTTCCGGCTCATG 2068
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      3860 gaagagtaacacgacgacgacgagagtgatccccagcagagtgagacactgtacatc 3919
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2069 GAGGAGCAAACTGCGCAGGCTAAATTTACGGCTCCACTGAATATAGACCGCGGTACATAC 2128
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      3920 ctggcctgattgagatgagatgaaatacaatggaatcccaagaacctgaaataaagg 3979
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2129 CAGGCTGGAGTGAAGGAGGGAATTGAAACCAACCCGCAAGACCTGAAATTTGAT---G 2185
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      3980 atgatgtcacacaagcaagctgtgttcacgacacatcctcgggagagccatctgtg 4039
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2186 AGGATACCTACAGACCACTTGGAGAGATTTCAAGCCGACACAGCATCCGAGGTGATTTTGG 2245
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4040 ctctcgtatgcccaaatgctcccatatcctcctcctccacctgaccaatcacagacc 4099
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2246 TATCACAAGTCCCAAGCCTTCCCTTGCTGACCAATACCAACCAAGTCAATACAGACC 2305
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4100 tgaagcgagaaatcaacgggggcagtcctcaatcgaacttgagacagcctccctgggagtg 4159
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2306 TTGATGCCACAGTTTCATGAGG---ATAAGATTATTTCTTACATGACACACAGAGATA 2362
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4160 attatgacccatgaaacggccacaaagatatcaltcgaataagtaagaatcattcttgatc 4219
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2363 ATTGTGATGTTGAAAGTTCAACGTTATATTCATAGCAATATAGTCAAGTATTTCTTATC 2422
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4220 tcagagacaagttcaatgaatcctcctcaatgtaatactactgctcctcctccaaaggag 4279
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2423 TAAGAGACAGTTTGAATGATGCTCTTCAAGTAATACTGATGATCTGTCAACCAAGAGAG 2482
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4280 ccaactctgaggaagctcttctttaaaccagaaacattactcttgaataagtcagag 4339
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2483 CCAACTCTCAAGGAAGCTTTTGATTTAAACAGCAAAATATCTCGAAGAAATGCAACCC 2542
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4340 atcttctcattgcatcagcgctgtgataagtcgactcgaatacagaataatcacaca 4399
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2543 ACATATTATTTGCGATTAAAGTATAGATTAAGCAATTGACATCAAAAAGTATCCACA 2602
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4400 tgcacagatcattctgttattctccacagactccgcacagaga---cacctagtcctg 4456
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2603 TTGCACAGATTAACCTTTGTTATCCCTCAACCAATCTGATGACATTGATCTACACCTA 2662
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Oy      4457 atgaacgtctgctctctgtcccaata 4483
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      2663 CTCCTACTCTACTCTACTCTCTGATA 2689
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 10
AC068071/c DNA HTG 28-MAY-2000
LOCUS AC068071 Homo sapiens chromosome 1 clone RP11-290N3, WORKING DRAFT SEQUENCE,
DEFINITION 15 unordered pieces.
ACCESSION AC068071
VERSION AC068071.4 GI:8099747
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS Aboia,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
FedeSPIEL,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,
Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 175591)
AUTHORS Aboia,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
FedeSPIEL,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
FedeSPIEL,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,

```

## TITLE

JOURNAL

## COMMENT

Mao,J., Marathe,R., Morehouse,A.J., Oefner,P., Palm,C.J.,  
 Ramirez,D., Wilhelmy,J., Yu,S. and Davis,R.W.  
 Direct Submission  
 Submitted (28-APR-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 On May 28, 2000 this sequence version replaced gi:7671256.  
 ----- Genome Center  
 Center: Stanford DNA Sequencing and Technology Development  
 Center

Center code: SDS/ND

Web site: <http://sequence-www.stanford.edu/group/human/>  
 Contact: [hum-info@sequence.stanford.edu](mailto:hum-info@sequence.stanford.edu)

Project Information  
 Center project name: 889

Center clone name: RP11-290N3  
 ----- Summary Statistics

Sequencing Vector: M13mp18; X02513  
 Chemistry: Dye-Primer; 08 of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319  
 Consensus quality: 160064 bases at least Q40

Consensus quality: 168051 bases at least Q30  
 Consensus quality: 170998 bases at least Q20

Insert size: 161152; agarose-fp  
 Insert size: 178431; sum-of-contigs

Quality coverage: 4.3x in Q20 bases; agarose-fp  
 Quality coverage: 3.9x in Q20 bases; sum-of-contigs.

\* NOTE: This is a 'working draft' sequence. It currently  
 consists of 15 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1013: contig of 1013 bp in length  
 \* 1014 1113: gap of unknown length  
 \* 1114 2493: contig of 1380 bp in length  
 \* 2494 2593: gap of unknown length  
 \* 2594 4311: contig of 1718 bp in length  
 \* 4312 4411: gap of unknown length  
 \* 4412 8684: contig of 4273 bp in length  
 \* 8685 8784: gap of unknown length  
 \* 8785 13051: contig of 4267 bp in length  
 \* 13052 22390: contig of 9239 bp in length  
 \* 22391 22480: gap of unknown length  
 \* 22491 34595: contig of 12105 bp in length  
 \* 34596 34695: gap of unknown length  
 \* 34696 46780: contig of 12085 bp in length  
 \* 46781 46880: gap of unknown length  
 \* 46881 59460: contig of 12580 bp in length  
 \* 59461 59560: gap of unknown length  
 \* 59561 74872: contig of 15312 bp in length  
 \* 74873 74972: gap of unknown length  
 \* 74973 89920: contig of 14948 bp in length  
 \* 89921 90020: gap of unknown length  
 \* 90021 107986: contig of 17976 bp in length  
 \* 107987 108096: gap of unknown length  
 \* 108097 129557: contig of 21461 bp in length  
 \* 129558 129657: gap of unknown length  
 \* 129658 150564: contig of 20907 bp in length  
 \* 150565 150664: gap of unknown length  
 \* 150665 175591: contig of 24927 bp in length.

## FEATURES

## source

1. 175591

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/clone="RP11-290N3"

/clone\_lib="RPCT human BAC library 11"

misc\_feature

1. 1013

```

misc_feature /note="assembly_name:Contig8"
1114..2493
/note="assembly_name:Contig10"
2594..4311
/note="assembly_name:Contig11"
4412..8684
/note="assembly_name:Contig12"
8785..13051
/note="assembly_name:Contig13"
13152..22390
/note="assembly_name:Contig14
clone_end:SP6
vector_side:left"
22491..34595
/note="assembly_name:Contig15"
34696..46780
/note="assembly_name:Contig16"
46881..59460
/note="assembly_name:Contig17"
59561..74872
/note="assembly_name:Contig18"
74973..89920
/note="assembly_name:Contig19"
90021..107996
/note="assembly_name:Contig20"
108097..129557
/note="assembly_name:Contig21"
129658..150564
/note="assembly_name:Contig22"
150665..175591
misc_feature /note="assembly_name:Contig23"
BASE COUNT 53213 a 32926 c 33415 g 54576 t 1461 others
ORIGIN

```

```

Query Match 23.2%; Score 1058.2; DB 2; Length 175591;
Best Local Similarity 95.3%; Pred. No. 4.6e-261;
Matches 1125; Conservative 0; Mismatches 2; Indels 54; Gaps 1;

QY 1 atgataaatctcctacacatacactctcccaagactaaccaggaagtgaatct 60
DB 62439 ATGGATTAATTCCTCAACACATACACTCTCCCAAGACTAAACAGAGAAAGTTGAATCT 62380

QY 61 ctgaataagccaataaaggagctctgatatgtggcaataatcaagaagctttaccaccaa 120
DB 62379 CTGAATTAACCAATTAACAGGCTCTGATATTGTGGCAATTAATCAAGAGCTTACCAACCA 62320

QY 121 aagagtcaggagaccagatgattcacaagctgaattctaccagaggtacaaggaggaactg 180
DB 62319 AAGAGTCCAGAGACCGATGATGATTCACAGCTGATTCACAGAGGTACAGAGAGAACTG 62260

QY 181 gtaccatctcctctgaaagtattacaatcaagaagaaggcaatcctcctaactcg 240
DB 62259 GTACCATTCCCTCTGAAGATTACAAATCAATGAAGAAAGAGCAATCCCTCCCTAATCTCG 62200

QY 241 tttaatagggccaacatctctgtatccaagaagcgggagagagacacaaccaaaaaagag 300
DB 62199 TTTTATGAGGCGCAACATATCTGTATACCAAAAGCGGGCAGAGACACAAACAAAAAGAG 62140

QY 301 aatttagaccaatactttagaacaattgagcaaaaatctcctaataaatactggca 360
DB 62139 AATTTTACACCAATATCTTTGTATGAACATGTGATGCAAAAATCCCTCAATTAATTAATCTG 62080

QY 361 aacgaataccagcagacatacaaaaagcttatccacatgatcaagtgggtccatccct 420
DB 62079 AACCGAATTCAGACACATCAAAAAAGCTTATCCACATGATCAAGTGGGCTTCACTCCCT 62020

QY 421 gggat-----a 426
DB 62019 GGGATGTAAGGCTGTCAATATACGCAATCAATAAATGTAATCCAGCATATAACAGCA 61960

QY 427 accaaagacaaaacacatgatattctcaatagatgcagaaaagccttligacaaatt 486

```

```

DB 61959 ACCAAAGACAAAACCATGATATCTCAATAGATGACAGAAAAGCCTTTGACAAAATT 61900
QY 487 caacacaccttatgtctaaaaacctcaataaattagatatgtatgggacatatctcaa 546
DB 61899 CAACACCCCTTCATCTTAAACCCCTCAATAAATTGATTTGATGGGACATATCTCAAA 61840
QY 547 ataaagaagctatctatggcaagccaagccaatcatatcatatctatgggaaaaactg 606
DB 61839 ATAAATAGAGCTATCTATGCGCAAAACCCAGCCATATCATATCTGAATGGGCAAAAAC 61780
QY 607 gaagcatcccttltgaaaacttggcacaagagagagatgcccctctctcaccatctatlc 666
DB 61779 GAAGCATTCCTCTTGAAAACCTGGCACAAGAGGATGGCCCTCTCCACCATCTCTATTTC 61720
QY 667 aacatagttttggaaagttctggccagggaacttaaggcagagaagaagaaatagggtttt 726
DB 61719 AACATAGTTTGGAGGTTCTGTGCGAGGGCAATTAAGCAGAGAAAGAAATAAAGGTTT 61660
QY 727 caattaggaagaagagagatcaaatgtgtccctgttgcaggtgagatgatgtatacccta 786
DB 61659 CAATTAGCAAAAAGAGAACTCAAAATTGTCCTGTTTGCAGGTGACATGATTTGTATACCTA 61600
QY 787 gaaaaccccatctctcagcccaaaatctcctttaaagctgataagaacttcagcaaaagtc 846
DB 61599 GAAAACCCCATCTCTCGAGCCCAAAATCTCTTAAGCTGATAGCAACTTCAGCAAAAGTC 61540
QY 847 tcaagatccaaaataatgaatgtacaaaatacagaagcttctataccataacagagaa 906
DB 61539 TCAGATATCAAAATCAATGATACAAAATCACAAGATTCCTTTACACCAATTAACAGAGAA 61480
QY 907 aagagagccaatatgaatgaatcccatccatcacatgtgtctcagaagaagaataaattc 966
DB 61479 ACAGAGAGCCAAATATGATGATTAATCACTCCCATTCACAAATTTGCTTCAAGAGAAATTAATTC 61420
QY 967 ctgaagatcccaacttacaagagatgtgaagagcctcttcaaggagaaactacaaacactg 1026
DB 61419 CTAGGAATTCACACTTACAAAGGATGTGAAGAGACCTCTTCAAGAGAACTTCACAAACACTG 61360
QY 1027 ctcaatgaataaagaagatacaaaaataaalgaaagaacatctcatgtcatgataaga 1086
DB 61359 CTCATATGAATTAAGAGAGATCAAAACAAATGGAAGAACATTCATGCTCATGATGATGCA 61300
QY 1087 agaatacatatcgtgaagaatggccatctactgtgcccaagatatt 1127
DB 61299 AGAATCAATATCTGTAATAATGGCCATTAAGTCCCAAGGTAAT 61259

```

```

RESULT 11
AL356270 140718 bp DNA HTG 23-JAN-2001
LOCUS Homo sapiens chromosome 1 clone RP5-873015, *** SEQUENCING IN
DEFINITION AL356270
ACCESSION AL356270.3 GI:9797470
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140718)
AUTHORS Pavitt,R.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Aug 12, 2000 this sequence version replaced gi:9213800.
COMMENT ----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dt873015

```

```
----- Summary Statistics -----
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 12% of reads
Sequencing vector: plasmid; L08752; 87% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 129686 bases at least Q40
Consensus quality: 134097 bases at least Q30
Consensus quality: 136445 bases at least Q20
Insert size: 138618; sum-of-contigs
Quality coverage: 3.10x in Q20 bases; sum-of-contigs quality
coverage: 3.21x in Q20 bases; agarose-1p

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 3292: contig of 3292 bp in length
* 3293 3392: gap of 100 bp
* 3393 10168: contig of 6776 bp in length
* 10169 10268: gap of 100 bp
* 10269 21700: contig of 11432 bp in length
* 21701 21800: gap of 100 bp
* 21801 29662: contig of 7862 bp in length
* 29663 29762: gap of 100 bp
* 29763 31862: contig of 2100 bp in length
* 31863 31962: gap of 100 bp
* 31963 36577: contig of 4615 bp in length
* 36578 36677: gap of 100 bp
* 36678 43943: contig of 7266 bp in length
* 43944 44043: gap of 100 bp
* 44044 52906: contig of 8863 bp in length
* 52907 53006: gap of 100 bp
* 53007 65117: contig of 12111 bp in length
* 65118 65217: gap of 100 bp
* 65218 70699: contig of 5482 bp in length
* 70700 70799: gap of 100 bp
* 70800 82141: contig of 11342 bp in length
* 82142 82241: gap of 100 bp
* 82242 87387: contig of 5146 bp in length
* 87388 87487: gap of 100 bp
* 87488 91496: contig of 4009 bp in length
* 91497 91596: gap of 100 bp
* 91597 104147: contig of 12551 bp in length
* 104148 104247: gap of 100 bp
* 104248 107008: contig of 2761 bp in length
* 107009 107108: gap of 100 bp
* 107109 110463: contig of 3355 bp in length
* 110464 110563: gap of 100 bp
* 110564 112670: contig of 2107 bp in length
* 112671 112770: gap of 100 bp
* 112771 116165: contig of 3395 bp in length
* 116166 116265: gap of 100 bp
* 116266 118389: contig of 2124 bp in length
* 118390 118489: gap of 100 bp
* 118490 129046: contig of 10557 bp in length
* 129047 129146: gap of 100 bp
* 129147 136534: contig of 7388 bp in length
* 136535 136634: gap of 100 bp
* 136635 140718: contig of 4084 bp in length.

FEATURES
source
1. 140718
/location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP5-873015"
/clone_lib="RPCT-5"
1. 3292
/note="assembly_fragment:00102
```

```
fragment_chain:1"
3393. 10168
/note="assembly_fragment:00109
fragment_chain:1"
10269. 21700
/note="assembly_fragment:00228
fragment_chain:2"
21801. 29662
/note="assembly_fragment:00045
fragment_chain:2"
29763. 31862
/note="assembly_fragment:00277
fragment_chain:3"
31963. 36577
/note="assembly_fragment:00849
fragment_chain:3"
36678. 43943
/note="assembly_fragment:00578
fragment_chain:4"
44044. 52906
/note="assembly_fragment:00490
fragment_chain:4"
53007. 65117
/note="assembly_fragment:01028
fragment_chain:5"
65218. 70699
/note="assembly_fragment:00015
fragment_chain:5"
70800. 82141
/note="assembly_fragment:00071"
82242. 87387
/note="assembly_fragment:00242"
87488. 91496
/note="assembly_fragment:00323"
91597. 104147
/note="assembly_fragment:00430"
104248. 107008
/note="assembly_fragment:00435"
107109. 110463
/note="assembly_fragment:00756"
110564. 112670
/note="assembly_fragment:00733"
112771. 116165
/note="assembly_fragment:00810"
116266. 118389
/note="assembly_fragment:00912"
118490. 129046
/note="assembly_fragment:01001"
129147. 136534
/note="assembly_fragment:01002"
136635. 140718
/note="assembly_fragment:01014"

BASE COUNT 44522 a 26618 c 25494 g 41976 t 2108 others
ORIGIN
Query Match 23.0%; Score 1051.8; DB 2; Length 140718;
Best Local Similarity 94.8%; Pred. No. 2e-259;
Matches 1120; Conservative 0; Mismatches 7; Indels 54; Gaps 1;

QY 1 atgataaattcctcaacatacaactctccaagactaaacagaagaagtgaattc 60
|||||
Db 105397 ATGATTAATTCTCGACACATFACACCTCCCAAGACTAATACGAGAGATTGAAATC 105456.

QY 61 ctgaatagaccataaagaagctctgatattggtgcaataatcaagagcttaccaccaa 120
|||||
Db 105457 CTGAATAGACCAATAACAGGCTCTGATTTGTGGCATAATCAAGACCTTACCAACCAA 105516

QY 121 aagaagtcacagacagtgatgattcaagtgaaattccacaagaggttacaagaggaact 180
|||||
Db 105517 AAGAGTCACGACACAGTGAATTACAGCTGAATTCTACCAAGAGGTACAGAGAACTG 105576

QY 181 gtaccattccctctgaagattacaatacaataaagaaggaatccctctaactcg 240
```

```

|||||
Db 105577 GTACATTCCTCTGAAGATATACATCAATAGAAAAAGAGACATCTCCCTTAATCG 105636
QY 241 ttatatggagcaacatcatctctgataccaaagccgggcaagacacacaaaaaagag 300
Db 105637 TTTTATGAGGGCAACATATCTCTGTATACAAAGCCGGGAGAGACACCAACAAAAAAG 105696
QY 301 aatttaagcaatactcttgatgaacatgatgcaaaatccctcaataaatactgca 360
Db 105667 AATTTAGACCAATATCTTTATGACATTTATGCAAAAAATCCTCAATTAATTAATCTGGCA 105756
QY 361 aaccgaatccagcagcacatcaaaaaagcttaaccacatgataagtyggcttcacct 420
Db 105757 AACGGAATCCAGCAGACATCAAAAAGCTTATCCACATGATCAAGTGCGCTTATCCCT 105816
QY 421 gggat-----a 426
Db 105817 GGGATGTAAAGCTGGTTCATATACGCAATCAATTAATGTAATCCAGTATATAAACAGA 105876
QY 427 accaaagaacaaacacatgatatctcaatagatgcagaagaagccttgacaaaatt 486
Db 105877 ACCAAAGCAAAAACCATATATTTCTCAATAGATGCAAAAAGCCTTTGACAAAATT 105936
QY 487 caacaacccctcatgctaaaaaacctcaataatagatagatgagacatatctcaaa 546
Db 105937 CAACAACCCCTCATGCTTAAAAACCTCAATTAATTAATGATGGAGCATATCTCAAA 105996
QY 547 ataataagactatctatgccaagaagcagaacacatcatcgaatgagggcaaaactg 606
Db 105997 ATAATAAGAGTATCTATGGAAGAACCCACAGCAATATCATATGTAATGGGCAAAAACGTG 106056
QY 607 gaacatccctcttgaaaacctggcacaagaagagatccctctctcacaactctatc 666
Db 106057 GAAGCATTCCTCTTGAAACTGCGCACAGACAGAGGATGCCCTCTCTCAACCTTATTC 106116
QY 667 aacatagtttggaaagtctggccagggcaattagcagaggaagaaataaagggttt 726
Db 106117 AACATAGTTTGAAGTTCTGGCCAGGCAATTAGCAGAGAAAGAAATTAAGGTTT 106176
QY 727 caatagaagaagaaggaagtaaatgtccctggttgcagtgagatgatactgataccta 786
Db 106177 CAATTAGGAAAAAGAGAGATCAAAATGTCCCTGTTGAGGAGACATGATGTATACCTA 106236
QY 787 gaaaaccccatctctcagcccaaaatcccttaagcgtgaagaagaattcagaagctc 846
Db 106237 GAAAACCCCATTTCTCGACCCCAAAATCTCTTAAGCTGATAGCAACTTCACCAAGTC 106296
QY 847 tcaggatacaaaatcaatgtaaaaaatcacaagaacatctctatacacaatacagaaga 906
Db 106297 TCAGATACAAATCAATGTACAAAATACACAGCATTCCTATACCAATTAACAGAGAA 106356
QY 907 acagagagacaaatcatgaaatgaactcccatcacaatgtctcagaagaataaataac 966
Db 106357 ACAGAGACCAAAATCATTAATGAAGTCCCATTCACAAATGTCTTCAAAAGAAATAAATAC 106416
QY 967 ctgaagaatccaaactagaaggaatgtaagagacccctcaagaagaactacaaacctg 1026
Db 106417 CTAGAATCCAACTTACAAAGGATGTGAAGACCTCTTCAMAGAGAACTACAAACACACG 106476
QY 1027 ctcaatgaataaaagagatatacaacaatggaagaacatctctatgtaagta 1086
Db 106477 CTCATGAATTAATAAGAGATGACAAACAATGGAAGAACATTCATGCTCATGATAGAGA 106536
QY 1087 agaatacaatcgtgaaatggccctactgccaagattat 1127
Db 106537 AGAATCAATATCTGTGAAAATGGCCACTACTGCCCAAGTAAT 106577

```

```

RESULT 12
LOCUS HS431C21 118426 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from PAC 431C21 on chromosome X contains ESTs
and ribonucleoprotein particle (hnRNP).

```

```

ACCESSION 282210
VERSION 282210.1 GI:1843448
KEYWORDS ribonucleoprotein particle; X.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 118426)
JOURNAL Mistry, S.
COMMENT Direct Submission
Submitted (15-JUN-1997) E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequests@sanger.ac.uk
On Feb 19, 1997 this sequence version replaced gi:1772965.
de Jong P.J., enquiries: http://bacpac.med.buffalo.edu/IMPORANT:
This sequence is the entire insert of clone 431C21. This sequence
has been finished according to sequence map criteria as follows.
An attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The true left end of clone 431C21 is at 1 in this sequence. The
true right end of clone 431C21 is at 118426.
431C21 is from the library RPO13 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR.pcrp2c.
FEATURES
Source
Location/Qualifiers
1..118426
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="X"
/clone="RP3-431C21"
/clone_1bp="RPI-3"
1..235
/note="AluSc repeat: matches 235. .1 of consensus;
incomplete repeat"
1217..1266
/note="2 copies of 25 mer 96 % conserved"
2293..2525
/note="L1M2 repeat: matches 1055. .815 of consensus"
2524..2817
/note="L1M1 repeat: matches 971. .664 of consensus"
2792..2950
/note="L1M1 repeat: matches 823. .984 of consensus"
2915..3337
/note="L1M9 repeat: matches 709. .277 of consensus"
3345..5663
/note="L1 repeat: matches 5274. .2925 of consensus"
5642..6044
/note="MSTR repeat: matches 1. .423 of consensus"
6049..6592
/note="L1 repeat: matches 2953. .2402 of consensus"
7337..7676
/note="MER25 repeat: matches 1783. .2136 of consensus"
7851..8054
/note="L1 repeat: matches 1791. .1985 of consensus"
8462..8491
/note="15 copies of 2 mer 87 % conserved"
<8919..>9526
/note="match: 3' EST N20961 clone 265803"
9004..9110
/note="AluSc repeat: matches 174. .299 of consensus;
incomplete repeat"
9566..9850
/note="MER43 repeat: matches 1. .268 of consensus"
complement(<10023..10126)
/note="match: 5' EST T63290 clone 80790"
10444..10506
/note="MER43 repeat: matches 210. .272 of consensus"
13487..13752
/note="AluX repeat: matches 3. .267 of consensus;
incomplete repeat"

```

```
repeat_region 13912..14264
/note="L1 repeat: matches 3591..3955 of consensus"
repeat_region 14260..14575
/note="L1 repeat: matches 4002..4324 of consensus"
repeat_region 14682..15269
/note="L1MA7 repeat: matches 596..1 of consensus"
repeat_region 15127..15415
/note="L1 repeat: matches 5390..5093 of consensus"
prim_transcript <15414..>15920
/note="match: 5' EST R19063 clone 33638; match: 5' EST
H21576 clone 160009"
repeat_region 16063..16270
/note="L1 repeat: matches 5083..4876 of consensus"
repeat_region 16327..16746
/note="MSTA repeat: matches 1..426 of consensus"
repeat_region 16747..17156
/note="MST-INTERNAL repeat: matches 1..413 of consensus"
repeat_region 17152..17493
/note="LTR7 repeat: matches 450..100 of consensus"
repeat_region 17410..17534
/note="LTR7 repeat: matches 131..1 of consensus"
prim_transcript <20771..21634
/note="match: multiple ESTs; match: 3' EST H60808 clone
208209; match: 5' EST H666214 clone 234021; Paired with
EST H666215 matching this clone; match: 3' EST H666215
clone 234021; Paired with EST H666214 matching this clone;
match: 5' EST H59570 clone 206711; Paired with EST H59571
matching this clone; match: 3' EST H59571 clone 206711;
Paired with EST H59570 matching this clone; match: 3' EST
H66951 clone 295823; match: EST T12379 clone A151"
prim_transcript complement(22040..22488)
/note="match: EST H60808 clone 28e05; similar to RTVL-H
related sequence"
repeat_region 22355..22757
/note="LTR7 repeat: matches 450..1 of consensus"
repeat_region 22766..23422
/note="MST-INTERNAL repeat: matches 408..1075 of
consensus"
repeat_region 23525..23740
/note="MSTD repeat: matches 1..216 of consensus"
repeat_region 23739..23847
/note="MSTC repeat: matches 224..402 of consensus"
repeat_region 24176..24211
/note="18 copies of 2 mer 89 & conserved"
repeat_region 24210..24284
/note="3 copies of 25 mer 83 & conserved"
repeat_region 24262..24297
/note="18 copies of 2 mer 83 & conserved"
repeat_region 24300..24337
/note="19 copies of 2 mer 84 & conserved"
repeat_region 24826..24853
/note="14 copies of 2 mer 96 & conserved"
repeat_region 24858..25456
/note="L1 repeat: matches 4213..3608 of consensus"
repeat_region 25471..26788
/note="L1 repeat: matches 3487..2133 of consensus"
repeat_region 26926..26977
/note="26 copies of 2 mer 83 & conserved"
repeat_region 28145..28279
/note="FLAM C repeat: matches 5..133 of consensus"
repeat_region 28711..28760
/note="L1MA2 repeat: matches 1005..1055 of consensus"
repeat_region 28813..28998
/note="L1PA15 repeat: matches 904..718 of consensus"
repeat_region 28996..29888
/note="L1PA2 repeat: matches 893..1 of consensus"
repeat_region 29739..34240
/note="L1 repeat: matches 5390..896 of consensus"
repeat_region 34236..35009
/note="L1 repeat: matches 772..8 of consensus"
repeat_region 35015..35737
/note="L1PA9 repeat: matches 741..1 of consensus"
repeat_region 35606..36916
```

```
/note="L1 repeat: matches 5390..4057 of consensus"
36911..38329
/note="L1 repeat: matches 3761..2341 of consensus"
repeat_region 38335..38586
/note="L1 repeat: matches 2230..1966 of consensus"
repeat_region 38587..38892
/note="L1US9 repeat: matches 303..1 of consensus"
repeat_region 38893..39448
/note="L1 repeat: matches 1977..1430 of consensus"
repeat_region 39243..39526
/note="MER25 repeat: matches 2136..1844 of consensus"
repeat_region 39618..40048
/note="L1PA7 repeat: matches 886..464 of consensus"
repeat_region 42147..42518
/note="MLT1B repeat: matches 11..374 of consensus"
repeat_region 42532..42565
/note="17 copies of 2 mer 88 & conserved"
repeat_region 44113..44162
/note="25 copies of 2 mer 96 & conserved"
repeat_region 44955..47064
/note="L1 repeat: matches 990..3058 of consensus"
repeat_region 47136..48331
/note="L1 repeat: matches 4176..5390 of consensus"
repeat_region 48187..49016
/note="L1MA4 repeat: matches 1..845 of consensus"
repeat_region 49038..49921
/note="L1PA14 repeat: matches 895..1 of consensus"
repeat_region 49772..50034
/note="L1 repeat: matches 5390..5129 of consensus"
repeat_region 50037..50318
/note="L1 repeat: matches 4864..5141 of consensus"
repeat_region 50300..50507
/note="L1MA5 repeat: matches 832..1045 of consensus"
repeat_region 50562..50780
/note="L1MA3 repeat: matches 231..3 of consensus"
repeat_region 51075..51929
/note="L1MA10 repeat: matches 1069..234 of consensus"
prim_transcript <52519..>53177
/note="match: multiple ESTs; match: 5' EST H87217 clone
220533; match: 5' EST R05992 clone 125327; match: 5' EST
R99371 clone 201201; match: 3' EST T64484 clone 80506;
match: 3' EST R42572 clone 30927; match: 3' EST R42487
clone 31053"
repeat_region 55349..55378
/note="15 copies of 2 mer 90 & conserved"
repeat_region 56474..56523
/note="25 copies of 2 mer 96 & conserved"
repeat_region 56550..56575
/note="13 copies of 2 mer 92 & conserved"
repeat_region 56780..57164
/note="MSTA repeat: matches 1..426 of consensus"

Query Match 23.0%; Score 1049.4; DB 9; Length 118426;
Best Local Similarity 80.9%; Pred. No. 8e-25;
Matches 1364; Conservative 0; Mismatches 166; Indels 155; Gaps 6;

QY 1 atgataaatctccacaacatacactctccaagaactaaacaggaaagttagct 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105431 ACGGATTAATTCTCGGACACATATACCGTCCCAAGACGAAGAGCAATCC 105430

QY 61 ctgaatagaccaataacaggctctgatattgtggcaataatacaagcttaccaccaa 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105491 CTGAATAGACCAATTAACAGTCTGAAGTGAAGCGCTATTAACGCCATCAACCAA 105550

QY 121 aaagatccaggaccagatgatcacaagctgaatttaccagggtacaaagggagacg 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105551 AAAAGTCCAGGACCAAGACCAATTCACAGCATTTCTACCAAGGATCAAGAGAGCTG 105610

QY 181 gtacattcccttgaagatattacaataagaagaaggaatcccttaactg 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105611 GTACCA-TCCCTTGTGAACATTTCCAAACATATGAAGAGAGGATCCCTCACTCA 105669

QY 241 ttattatgagccaacalcctgatccaagaacgycggcagagacacacaaacaaagag 300
```

Dh	105670	TTTTATGAGGGCAGACATCATCCTGATATCCCAAACTGGCAGAGCACAACAAAAAGAA	105729
Oy	301	aattttagaaccaatattcttgaagaacatgatgacaaatctcctaataaatcttgcga	360
Dh	105730	AATTTCAGTGTCATATCCCTCATGAAATGCATGCATGCAAAAATCCCATATAAATCTGGCA	105789
Oy	361	aaccgaatccgcagcagacatcaaaaagcttatccaccatgatcaagtggtcttaccct	420
Dh	105790	AACCAAAATCGACGACATCAAAAAAGCTTTATACATACGATCAAGTTGGCTTCATCCCT	105849
Oy	421	gggat-----a	426
Dh	105850	GGGATGCCAACTGGTTCAACAAATGTATAATTAATGATGATCCATCACAATAACAGA	105909
Oy	427	accaaagaacaaaacacatgatattatctcgaatagatgcagaanaagccttgaacaaat	486
Dh	105910	ACCATTAAACAAAACCAACAAATATATCTCATPATATGCAGAAAAGGCTTCAACAGATT	105966
Oy	487	caacgaaccttcctatgcataaaaacctcaataatagatatgtatgagacatatctcaa	546
Dh	105970	CAACGCCCTTCATGCTTAAAAACCTCATATTAAGTATCAATCAATGAAAGTATCTCTAAA	106029
Oy	547	ataataagagctatctatgycagaagccacagccaatctcatctgaaatgggcaaaaactg	606
Dh	106030	ATAATTAAGCTATTATTATGACAAAACCCAGCCAAATTCATCTGAAAAGGCAAAAACCTG	106088
Oy	607	gaagca-ctccctcttgaanaacttgcacagaagaacaggaatgcctcttcacacaactctat	665
Dh	106090	GAGCAGATCCCTTGATTAATCTGGCACA-----GGATGCCCTCTGCACACACTCTATT	106144
Oy	666	caacatagttcttggaagttctgcgccaggcgcaatgagcagagagaagaataaaggctt	725
Dh	106145	CAACCTTACTATTGGAAAGTTCTGGCCAGGCAATAGGAAAGAGCAAAATTAAGGGCTAT	106204
Oy	726	tcaattagaaagaagagagatcgaatgtccctgtttgcaggtgacatggattgatacct	785
Dh	106205	TCAATTAAGGAAGAGNGAGATCAAAATCGCTCTGTGTTGGAGATGACATGATTGTATTTT	106264
Oy	786	agaaaaccccatctctcagcccaaaatctccttaagctgataagcaactctcaagcaat	845
Dh	106265	AGAAAACCCCATTCGCTCAGGCCCAAAATCTCTTAAGCTGATAGCAAAATTCACACAAGT	106322
Oy	846	ctcagatatacaaatcaatgtaacaaaatcaacaagcatctccatacacaataacagaga	905
Dh	106325	CTCAGATTAACAAAATCAATGTGCAAAAATTCACAAGCAATCTCTTAATACCAATTAACAGACA	106384
Oy	906	aacagaagagccaatcatgaaatgaaatcccatcaaatgcttcaagaagaataaata	965
Dh	106385	AACAGAGGCCAAATCATGATGAAATGCCATTCGATTAATTCCTTCAAAAGAAATTAATA	106444
Oy	966	cttagaaatccaacttacaagggatgtgaagagacctcttcaagagagaactacaaccact	1025
Dh	106445	CCTATGATTAACACTTACAAG-----GAAGACCTCTTAAGAGCAACATCAAAACACT	106499
Oy	1026	gtccaaatgaataaagaaggaatacaaacaaatggaagaacattccatgctcatgataagg	1085
Dh	106500	GCACAGAGAAATTAAGAGGAGCACAAACAAATGCAAAAAGATTTCATGCTCATGAATATG	106558
Oy	1086	aagaaatcaatctgtgaanaatggtccatattgcccagaatattgctagatataaaggat	1145
Dh	106560	AAGATTCATATTGTGAAAATAGGCCATATCTCCCAAGATTAATTAACCAATTCATTCCTAT	106619
Oy	1146	-----	1145
Dh	106620	CTCCATCAAGCTACACATGACTTTCTTCACAGAACTGGAAAAAACTACTTTAAAGTTCA	106679
Oy	1146	-----tcaattagaaagaaggaatcaaatgttcc	1176
Dh	106680	ATGACCCCAAAAAGAGCCCGCATATAGCCAAAGACAATTCATAAGCAAAATGTCTCT	106739
Oy	1177	ctgtttgaagtgcacatgattgtatatcttagaaaaaccccatgttctcagcccaaatctc	1236

```

Db 106740 CMTTGGAGATGACATGATGTATATATTAGAAAACCCAGGCTCTCAACTCAATCTT 106799
OY 1237 cttaagctgataagcaacttcagcaaaagctcagatatacaaaatcaatgataaata 1296
Db 106800 CTTAAGCTGATTAAGAAACTTCTAGCAAAAGTCTCAGATACCAATCAATGTGCAAAATATA 106859
OY 1297 caagatctctatacaccacacacagacaaaagaagagcaaatcatctgtgaactccca 1356
Db 106860 CAAAGATTCTTATACACCAATATATATAACAAACAGAAAGCCAAATCTGAGTGAATCCCA 106919
OY 1357 ttcaaatctgtctccaaaagaaataaatactcaggaatccaaactcaagaggaactgaag 1416
Db 106920 TTACAAATTTGCTACAAAGAGATTAATAATCCCTATGATATCACTTCAAGAGATGTGAAG 106979
OY 1417 gacctcttcaagagagaactactacaaccactgctcctaaggaataaaagaggatcaaaa 1476
Db 106980 GACCTCTTCAAGAGAAACTACAAACACACGCTCAACGAATATAGAGAGAAACAAACAAA 107039
OY 1477 tggaaagaacatttccttcgatgataggaagaatcaatcatgttgaaatggtccatagc 1536
Db 107040 TGGAAAGACATTTCATGCTCATGATGAGTAGAGAAATCAATATGATGTGAACAGCGCATATCTG 107099
OY 1537 cccaa 1541
Db 107100 CCCAA 107104

RESULT 13
AC010275/3
LOCUS AC010275/c
DEFINITION Homo sapiens chromosome 5 clone CTC-499J9, WORKING DRAFT SEQUENCE,
15 ordered pieces.
ACCESSION AC010275
VERSION AC010275.5 GI:9256179
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 196805)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 196805)
DOE Joint Genome Institute.
Direct Submission
Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7710793.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 377021
Center clone name: CIT-HSPC_499J9
-----
Summary Statistics
Consensus quality: 185697 bases at least Q40
Consensus quality: 193135 bases at least Q30
Consensus quality: 194420 bases at least Q20
Estimated insert size: 194000; pulse field gel estimation
Estimated insert size: 196000; sun-of-contigs estimation
Quality coverage: 6.0 in Q20 bases; pulse field gel estimation
Quality coverage: 5.93 in Q20 bases; sun-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitor.
* This sequence will be replaced

```

```
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 10284: contig of 10284 bp in length
10285 10384: gap of unknown length
10385 23853: contig of 13469 bp in length
23854 23953: gap of unknown length
32954 35219: contig of 11265 bp in length
35219 35319: gap of unknown length
80205 80205: contig of 44887 bp in length
80306 84921: contig of 4616 bp in length
84922 85021: gap of unknown length
101022 101022: contig of 16001 bp in length
101023 101122: gap of unknown length
101123 106588: contig of 5466 bp in length
106589 106688: gap of unknown length
106689 113556: contig of 6868 bp in length
113557 113656: gap of unknown length
113657 117493: contig of 3836 bp in length
117493 117592: gap of unknown length
130679 130778: contig of 13086 bp in length
130779 132182: contig of 1404 bp in length
132183 132282: gap of unknown length
132283 136437: contig of 4155 bp in length
136438 136537: gap of unknown length
136538 154086: contig of 17549 bp in length
154087 154186: gap of unknown length
154187 182675: contig of 28489 bp in length
182676 182775: gap of unknown length
182776 196805: contig of 14030 bp in length.
FEATURES
source
1. 196805
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTC-493J9"
BASE COUNT 62003 a 36870 c 35147 g 61383 t 1402 others
ORIGIN
Query Match 22.6% Score 1033.8; DB 2: Length 196805;
Best Local Similarity 77.5%; Pred. No. 8.7e-255;
Matches 1365; Conservative 0; Mismatches 174; Indels 223; Gaps 5;
QY 1 atgataaattcctcaacacatacaactctcccaagactaaacccgagagaagtgaatct 60
Db 137593 ATGGATACATTCTCGTCGACATACACTCTCCCAAGACTAAACCGAAGAGATTGATCT 137534
QY 61 ctgaatagaccaataacagagctctgatatgtgccaataacaaagacttaaccaacaa 120
Db 137533 CTGAATAGACCAATACAGAGCTCTGAAATTTGTGCAATTAATCAATGTTTACCAACCAA 137474
QY 121 aagaagtcagagacagatgattacacagctgaaattctcaacagagtgtaacaggagaaact 180
Db 137473 AAGAGTCAGAGACAGATGATTCACAGCCGAATTCACCAAGAGATCAAGAGAGAACTG 137414
QY 181 gtaacattccctctgaagaatcttaacataagaaaagagaaacccctccactactg 240
Db 137413 GTACCATCTCTTCTGTAACATATTCATATCAATGAAAAAGAGGAATCTCTCCCTCACTCA 137354
QY 241 tttaataagagccaacatcatctctgataccaagccggcgagagacacacccaanaaagag 300
Db 137353 TTTTATGACACCAATATCTTGATGAAACATTGATGCAAAAAATCTCATTAATAAATCTGGCA 137294
QY 301 aatttagaaccaataatcttgatgaacatgtagcaaaaatccctcaataaatactgca 360
Db 137293 AATTTTAGACCAATATCTTGATGAAACATTGATGCAAAAAATCTCATTAATAAATCTGGCA 137234
QY 361 aaccgaatcagagcacatcaaaaagcttatccacatgatcaaggggttcacacct 420
Db 137233 AACCGAATCAGACGACATCAAAAGCTTATCCACCATATCAAGTGGGCTTCATCTCT 137174
```

```
QY 421 gggat-----a 426
137173 GGGATGCAAGGCTGCTTCATATATGCAATATCAATAATGTAATCCAGCATATAACAGA 137114
QY 427 accaaagacaaaacccaatgatataatcgaatgacgaaagaccccttgcaaaatc 466
Db 137113 GCCAAAGCAAAAACCATGATATATCTCAATATATCAAAAAAGCCTTGGACAAAATT 137054
QY 487 caaacacccctcatgctcaaaaacccctcaataaattagatatgtagggacatctcaaa 546
Db 137053 CAACAACCCCTTACGCTAAAAACCTCAATTAATATAGGTATGAGGAGCTATTTCAGA 136994
QY 547 ataataagctatctatgycnaaagccacagccaatalcatactgaatlyggcaaaaatg 606
Db 136993 ATAATAAGACTATCTATGACAAACCCACAGCCAAATATCATAGATGGGCAAAAACCTG 136934
QY 607 gaagcatctcccttgaaaactgycacaaagcagagatgacccctctcaccacccatctc 666
Db 136933 GAAGCATTCCTTTGAAACTGGCACAAAGACAGGATGCCCTCTCTACCGCTCTATTTC 136874
QY 667 aacatagttttggaagttctgccaagggcaattagcagagaagaagaataaagggttt 726
Db 136873 AACATAGTGTGGAGATTCTTGCCAGGCAATCAGCGAGAGAAAGAAATTAAGGTATT 136814
QY 727 caattagaagaagaggaagtcnaattgctccgttctgcaagtacatgattgataccta 766
Db 136813 CAATTAGGAAGAAGAGAAAGTCAAAATTTGCTCCGTTCAGACGACATGATGTTTATCTA 136754
QY 787 gaaaccccatctctcgaagcccaaaatctccttaagctgtgtaagcaatctcagcaaatc 846
Db 136753 GAAACCCCATCTCTGAGCCCAAAATCTCTTAAGCTGATGAAGCACTTCAGCAAAAGTTC 136694
QY 847 tcagatcacaaaatcaatgtaacaaaatcaacagacatctccatataccaaataacagaa 906
Db 136693 TCAGATACAAAATCATATGTATGCAAAAATTCACAGCATTTCTTATACCAACACAGCA 136634
QY 907 acagagagccaatcatgaaatgaaatcccatcaacaatctgctcnaaagagaataaaatc 966
Db 136633 ACAGAGAGCCAAATCATGATGTAAGTCCCATTCACATTCCTTCAAAGAAATTAATAATAC 136574
QY 967 cttaggaatccaacttacaagatgtaaggaactctcaagagaactcaaacactc 1026
Db 136573 CTAGGAATCCAACTTAC-AGGAGTGTAGAGACCTCTNNNNNNNNNNNNNNNNNNNNNN 136515
QY 1027 ctaatgaataaagaagatacaacaatgagaagacatactcatgctcatg----- 1080
Db 136514 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136455
QY 1081 -----atagaagaatacaatcgtga 1102
Db 136454 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136395
QY 1103 aaatggccatcactgcccgaataatgtagat----- 1134
Db 136394 CCAATGCCCTATGTCATATCTGAATGTGCGAATCTCGAAGCCCTTCCTATGAAATACGTA 136335
QY 1135 ----- 1134
Db 136334 TAAGACAAGATGCCCTTTCAACACTTCTATTCAAAATATGATTAACAAAGTTCTGGCCA 136275
QY 1135 -----ataaaggatcaatcattagaaaga 1160
Db 136274 CAGAAAGAAATAGGCAATCAGGCACAGAAAGAAATTAAGGCAATTAAGAGAGACA 136215
QY 1161 ggaagtcnaattctccgttctgagatgacatgattgatactctgaaagaccccatgt 1220
Db 136214 GGAAGTCAAAATTTGTTTGTTCAGATGACATGTTATATTTAGAAACCCCATGTG 136155
QY 1221 ctaagcccaaatctccttaagctgataaagcaactcgaagaagctcgaagatacaaat 1280
Db 136154 CTCAGCTCAACACACCTTAAAGCTGATTAAGCAACTTAAGCAAAAGTCTCAAGGTACAAAAT 136095
```



\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of 'N', but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
misc_feature      1. .1215
                  /note="assembly_name:Contig10
                  clone_end:T7
                  vector_side:left"
```

```
misc_feature 1316. 2643 /note="assembly_name:Contig11"
misc_feature 2744. 4582 /note="assembly_name:Contig12"
misc_feature 4683. 6572 /note="assembly_name:Contig13"
misc_feature 6673. 8059 /note="assembly_name:Contig14"
misc_feature 8160. 9839 /note="assembly_name:Contig15"
misc_feature 9940. 11957 /note="assembly_name:Contig16
clone_end:SP6
vector_side:right"
misc_feature 12058. 14675 /note="assembly_name:Contig17"
misc_feature 14776. 16933 /note="assembly_name:Contig18"
misc_feature 17034. 19985 /note="assembly_name:Contig19"
misc_feature 20086. 24485 /note="assembly_name:Contig20"
misc_feature 24386. 28286 /note="assembly_name:Contig21"
misc_feature 28387. 32573 /note="assembly_name:Contig22"
misc_feature 32674. 36708 /note="assembly_name:Contig23"
misc_feature 36809. 41692 /note="assembly_name:Contig24"
misc_feature 41793. 47698 /note="assembly_name:Contig25"
misc_feature 47799. 54305 /note="assembly_name:Contig26"
misc_feature 54406. 61109 /note="assembly_name:Contig27"
misc_feature 61210. 68598 /note="assembly_name:Contig28"
misc_feature 68699. 77839 /note="assembly_name:Contig29"
misc_feature 77940. 85716 /note="assembly_name:Contig30"
misc_feature 85817. 93966 /note="assembly_name:Contig31"
misc_feature 94067. 103047 /note="assembly_name:Contig32"
misc_feature 103148. 113126 /note="assembly_name:Contig33"
misc_feature 113227. 122160 /note="assembly_name:Contig34"
misc_feature 122261. 135587 /note="assembly_name:Contig35"
misc_feature 135688. 147635 /note="assembly_name:Contig36"
misc_feature 147736. 162548 /note="assembly_name:Contig37"
misc_feature 162649. 163761 /note="assembly_name:Contig4"
misc_feature 163862. 165252 /note="assembly_name:Contig7"
misc_feature 165353. 166804 /note="assembly_name:Contig8"
misc_feature 166905. 168185 /note="assembly_name:Contig9"
BASE COUNT 50525 a 30701 c 29699 g 54153 t 3107 others
ORIGIN
```

Query Match	22.4%	Score 1024.8;	DB 2;	Length 168185;
Best Local Similarity	80.8%;	Pred. No. 1.8e-252;		
Matches 1362; Conservative	0;	Mismatches 167;	Indels 157;	Gaps 8;
QY 1 atggaataattcctcaacacatacacctctcccaagactaacaccaggaagaagttaattc	60			

Db 124416	ACGGATTAATTTCTGGACACATATACCGTCCCAAGACGAAACAGAGAAAGTCATAATTC	124475
Qy 61	ctgaatagaccatatacagctctgatatctgtgacaaatlaacaaagcttaccaccaa	120
Db 124476	CTGAAATAGACCAATATACAAAGTTCTGAAATTGAGGCCGTAAATTGACACCTACCAACCA	124535
Qy 121	aagagtcacagacacagatgatctcaacagctgaattctaccagagttacaaaggaactg	180
Db 124536	AAAACCTCCAGGACCCAGACAGATTTACACACCCATTTTCTACAGAGAGGTCAAAAGAGAGCTG	124595
Qy 181	gtaccattccctctgaagatgttcaatcaatataaaaaagaggaatccctccctaactcg	240
Db 124596	GTACCA-TCTCTTCTGAAACTATTCGAAAACATATAGAGAAAGG-AACTCTCCCTTAACCTCA	124653
Qy 241	ttctatagagccaacatcatcctctgatatatacgaagccgggcagagacacaaacaaanaag	300
Db 124654	TTTTATGAGGCCAGATCATCTGTATACCCAAACCTGGAGAGACACACAAAAAAGAA	124713
Qy 301	aatttagacccaatactcttgaatgaacttgaatgcaaaaatctctcaataaatacttgca	360
Db 124714	AATTTCAAGTCAATTTCCCTGATACATTCATGTCGAAAAATTCCTCATATAAATACTGGCA	124773
Qy 361	aaccgaatccagccagcaatcaataaaagcttccacatgatcaagtyggcttcatccct	420
Db 124774	AACCAAAATCCAGCACACATCAAAAAAGCTTATACATCAGATCAAGTTGGCTTCATCCCT	124833
Qy 421	ggat-----a	426
Db 124834	GGGATTCGCAACCTGGTTCAACAATATGTAATTAATTAATTAATTCATACATTAACACGA	124893
Qy 427	accaaagaacaaacacacatgatatactcaatagaatgcagaanaagccttgaacaaat	486
Db 124894	ACCAATTAACAAAAACCAACAAATTTATCTCATATAGATACAGAAAAAGGCTTCAACAGATT	124953
Qy 487	caacaaccttcataatgctataaaaccttaataataatgatataatgttgagacatatccaa	546
Db 124954	CAACGCCCTTCATGCTAAAAAAGCTCAATTAACATTAATTAATTAATTAATTAATTAATTAAT	125013
Qy 547	ataataagagctatctatgycanaagccacagccacatactatctgaaatgycgcaaaaactg	606
Db 125014	ATAATTAAGAGCTATTTATGACAAACCCAGCCCAATTCATATACGAAAGGGCAAAAAATG	125073
Qy 607	gaagca-ttcccttgaactgycacaaagagagatgcctctctaccactctat	665
Db 125074	GAAAGCAGATCCCTTTGATTAATCGGCACAA-----GGAATGCCCTCTCACACACTCTATT	125128
Qy 666	caacatgatttggaaagttctgcccagggcaatttaggcagagagaagaataaagggtt	725
Db 125129	CAACTATGATTTTGGAAAGTTCTGGCCAGGCGAATAGAGAAAGAGAAATAAAGGGTAT	125188
Qy 726	tcaattagaaagagggagatcgaatttgtccctgtttgcaggtacatactgtatccct	785
Db 125189	TCAATTAGGAAGAGAGAGAGATGCAAAATGCCCTCTCTTTGCAATGACATGACTGTATTTT	125248
Qy 786	agaaaaaccatctctcagcccaaaatctcttaagctgataagcaacttcagcaagt	845
Db 125249	AGAAAACCCCATCTGCTCTCAGCCCAAAATTCCTTAAGCTGATTAAGCAAAATTCAGCAAGT	125308
Qy 846	ctcaggtataaaaatcaatgtatacaaaaatccaagatctctataacccaataacagaga	905
Db 125309	CTCAGGATTAACAAATTCATATGTCGAAAAATTCACAAAGCTTCCTTAACCCAAATTAACGACA	125368
Qy 906	aacagagagccaatcatgaatgaactcccatctcaaatctgcttcaagaagaaataaata	965
Db 125369	AACAGAGAGCCAAATTCATGCTGATGACTCCCATTTCTGTAATTGCTTCAAAAGGAATTAATAA	125428
Qy 966	cctaagaaatccaacttcaaaaggaatgtgaaggaactcttcaagagaaacttacaacact	1025
Db 125429	CCTAGGAATTAACAATTCACAAAG-----GAAGGAGCTCTTAAGGACAACTACAAACCACT	125483
Qy 1026	gtcctaatgaataaagaaggaatacaacaaatggaagaaacttccatgtctcatgtatagg	1085



Query Match	22.0%	Score 1005.4;	DB 2;	Length 149428;
Best Local Similarity	92.4%;	Pred. No. 1.7e-247;		
Matches 1091; Conservative	0;	Mismatches 36;	Indels 54;	Gaps 1;

Oy	1	atgggtaaatcttcctcaaacatctacacgtctcccaagacatcaaccaggaagaattgaatc	60
Db	27266	ATGGATTAATTTCTCTGCACACATACACCCCTCCCAAGACTTAACGAGGAGAACTTGACTCT	27207
Oy	61	ctgaatagaaccaatacaagcctctgatatctgycgaataatcaagagcttaccaccaa	120
Db	27206	CTGAATAGACCAATTAACAGGCTCTGAAATTTGGCCAATTAATAGCTTACCAACCAAA	27147
Oy	121	aagagtcacagaccagatfgatctcaagctgtaattctaccagagttacaaggaagactg	180
Db	27146	AAGAGTCCAGGACCAAGATGATTCCACGCCAAATTTCTACCAAGAGTACAGAGGAACCTG	27087
Oy	181	gtaccatcccccctgaagtatcaatcaatagaaaagagcaatccctccctaactg	240
Db	27086	GTACCATTCCTCTGAAACTATTCCATCAATAGAAAAAGAGGAAATCTCCCTTAATCTCA	27027
Oy	241	tttatagagccaacatcaatccctgatalaccaaaagccgycgagagacaacccaataaagag	300
Db	27026	TTTTTATGAGGCGCAGATCATCTCTGATACCAAGCCAGCGACAGACACACCAAAAAAGAG	26967
Oy	301	aatttaagccaatctcttfgtgaacttfgatgcaaaaaatccctcaataaatactatgga	360
Db	26966	AATTTTAGACCAATATCTTGATGTAACCTGATGCAAAAATCTCTCAATTAATTAATCTGGCA	26907
Oy	361	aaccgaatccagcagcagacatcaaaaaagcttaccacatgatacaagtgygcttcaacct	420
Db	26906	AACCGAATCCAGCAGCACATCAAAAAAGCTTATCCACATGATCAAGTGCGCTTATCCCT	26847
Oy	421	gggat-----a	426
Db	26846	GGGATGCAAGGCTGGTTCATATATATGCAAAATCAATTAATGTAATCCAGATATAGACAGA	26787
Oy	427	accataagacaaaacacacatgatatcccaatagatgcagaaaagccttgcacaaatt	486
Db	26786	ACCAATAGCAAAACACACATGATTTATCTCAATAGATGCAGAAAAGCCTTTACCAAAATT	26727
Oy	487	caacaacccttcatagtctaaaaaaccttcaataatagatatatgaltgycgacatatctcaa	546
Db	26726	CAACAAACCCTTCATGCTAAAACTCTCAATTAATTAATGATATGATGAGGAGCGATATCTCAA	26667
Oy	547	ataataaagctatctatgycgaagccacagccaatcatcatctgaattgycgaaaaactg	606
Db	26666	ATAATTAAGAGCTATCTATGACAAACCCACAGCCAATATCATATCTAAATGGGCAAAAAACCTG	26607
Oy	607	gaagcatccctcttgaaaccttgcacaagaacagagatgcccctcttcacacatccctatctc	666
Db	26606	GAGCATCTCCCTTGAAAACTGGCACAAGACAGGATGCCCTCTCTCACACATCCTATATTC	26547
Oy	667	aacatagcttctggaaagctctgycgagggcaattlagccaggaaggaataaagggtcttc	726
Db	26546	AACATAGTGTGGAAAGTCTCTGGCCAGGGCCAATATAGCGAGGAGAAAGAAATTAAGGGTATT	26487
Oy	727	caattaggaanaaggaagatcaaatctgfcgccgtttgacagygaaatgatgttataacta	786
Db	26486	CAATTAGGAANAAGAGAGTCAAAATTTGCCCTGTTGGCAGATGAATGATATGTATATCTCA	26427
Oy	787	gaaaaccccatctctcagccccaataatcccttlaagctgataagacaacttcagcaaaagtc	846
Db	26426	GAAAAACCCCATTTGTCTCAGCGGCAAAATCTCCTTAAGCTGATAGGCACACTTCCAGCAAACTC	26367
Oy	847	tcaggtatcaaaaatcaatgtatacaaaaatcaacaagcattccctatatcaccaataacagaga	906
Db	26366	TCAGGTATCAAAAATCAATGTACAAAAATCACAAAGCATTTCTTATACCACAATTAACGACAAA	26307
Oy	907	acagagagccaatcaatgaatgaactcccatccaatctgcttcaaaaaagataataaac	966
Db	26306	ACAGAGACCAATCATATAGTGAACCTCCCATTTCACAATTTGCTTCAAAAGATATAAATAATTC	26247

QY	667	cttggagatcccaattcaaggagatcttggaaggaccctttcaaggagaaactcaaacacacty	1026
Db	26246	CTAGGAAATCCAACTTATCAAGGGATGTGAAGGACCTCTTCAAGGAGACCTTCAAAACCACTG	26187
QY	1027	ctcaatggaataaaggagatacaacaacttggagaacattcatctgatatgagga	1086
Db	26186	CTCAATGTAAATTAAGAAGGATTCAAACAAATGGAGAGACATTCATGCTCATGCTGATGAGGA	26127
QY	1087	agaatcaatctcgtgnaaatgacaaactgacacgaattat	1127
Db	26126	AGAAATCAATATCTGTGAAATGGCCATCTACTGCCCATGTGTAAT	26086

```
Search completed: March 30, 2002, 13:24:16
Job time: 22938 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:32:04 ; Search time 5438.31 Seconds  
(without alignments)  
6334.866 Million cell updates/sec

Title: US-09-867-034-4

Perfect score: 3206

Sequence: 1 ttgcgctcagatgtaaacact.....ttgttagatgtaaaaaaaa 3206

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estsba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hlc:\*  
10: gb\_estcl:\*  
11: gb\_estc2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	401	12.5	582	10	AM939544 QV1-DR007
2	401	12.5	582	10	AM939591 QV1-DR007
3	343.6	10.7	533	11	BE871265 601446959
4	343	10.7	517	11	BE871265 601446959
5	325	10.1	536	10	BE871265 602464172
6	311.2	9.7	546	11	BE866879 601442694
7	301.8	9.4	461	10	BE866879 601442694
8	298.8	9.3	807	11	BE870652 601447779
9	286.6	8.9	468	10	AA315623 EST187445
10	272.4	8.5	485	10	AA633302 nq58e03.s
11	260.6	8.1	477	10	AW751697 I14-CT008
12	252.8	7.9	454	10	AT732655 nm99912.x

C 13	250.4	7.8	449	10	AT573096
C 14	246.6	7.7	531	10	AM363038
C 15	243.6	7.6	431	10	AA581944
C 16	243.6	7.6	498	10	AM274215
C 17	242	7.5	432	10	AM191851
C 18	239.2	7.5	449	10	AM842604
C 19	229.6	7.2	484	10	AT1821845
C 20	228.8	7.1	427	10	AA053660
C 21	226.6	7.1	460	10	AM842603
C 22	222.8	6.9	570	10	AM363009
C 23	220	6.9	500	10	AM363039
C 24	220	6.9	566	10	AM363010
C 25	219.4	6.8	396	10	AT180846
C 26	219.2	6.8	580	10	AA877900
C 27	218.8	6.8	411	10	AT127359
C 28	216.4	6.7	325	10	AM752537
C 29	214.6	6.7	381	10	AM293665
C 30	213.2	6.7	342	10	AA582019
C 31	212.2	6.6	378	10	AT1620830
C 32	211.6	6.6	315	10	AM843332
C 33	204.2	6.4	355	10	AA366856
C 34	202	6.3	428	10	AM080947
C 35	194.8	6.1	563	10	AT1431674
C 36	192.6	6.0	272	10	AA601073
C 37	191.6	6.0	586	10	AM374894
C 38	190.8	6.0	397	10	AM939557
C 39	190.8	6.0	397	10	AM939557
C 40	188.2	5.9	369	10	AM939880
C 41	185	5.8	381	10	AT1278909
C 42	169.8	5.3	352	10	AM191847
C 43	169.8	5.3	586	10	AM603295
C 44	164.6	5.1	238	11	T24571
C 45	155.4	4.8	223	10	AT708873

#### ALIGNMENTS

RESULT 1  
LOCUS AM939544 582 bp mRNA  
DEFINITION QV1-DR0072-110200-066-a10 DT0072 Homo sapiens EST  
ACCESSION AM939544  
VERSION AM939544.1 GI:8114990  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=at2-QV1-DR0072-110  
200-066-a10&t3=2000-02-11&t4=1)  
Seq primer: puc 18 forward



QY	750	ccctgtgagaagaataatcgattatgcttctgttaagcttgcataatacaagatllttgcctgt	809
Db	362	CCCTGTGAGAGAAATATCGTTATGTGTAAAGCTGC-----ATAATACAACTTTTGCCTGT	416
QY	810	gtttgaagggatctactacaactcttctaacgtgtaagaagaagagatctccctggag	869
Db	417	GTTTGAAGGAGGATTTACTACAACTTTCTTACTATGTATGAGAAAGAAAGATTTCCCTGG-G	475
QY	870	aagatttcagttacagatlcagaacacatttgacccacagaagagaacatccatgcctat	929
Db	476	AAGATTTCAGTGCAGATATCAGAAACATTGTGACCCAGAAAGAAACATTTCATGCGCTAT	535
QY	930	caagactgcacatgtgaataactactagcttgttttaagaatgtatttg	976
Db	536	CAAGACTTGCATATGTAAGAAATTTACTAGCTTGTTTAAAGATGATTTTG	582
RESULT 3			
LOCUS	BE871265	533 bp	EST
DEFINITION	601448959p1 NIH_MGC_65	Homo sapiens	CDNA clone IMAGE:3852686 5',
ACCESSION	BE871265	BE871265	BE871265
VERSION	BE871265.1	GI:10320041	EST.
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strusberg, Ph.D.		
	Email: <a href="mailto:cgapbs-rt@mail.nih.gov">cgapbs-rt@mail.nih.gov</a>		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	<a href="http://image.llnl.gov">http://image.llnl.gov</a>		
	Plate: LLM9575	row: f	column: 15
	High quality sequence stop: 457.		
FEATURES	Location/Qualifiers		
source	1..533		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone IMAGE:3852686"		
	/clone_id="NIH_MGC_65"		
	/tissue_type="adenocarcinoma"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;		
	Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.		
	Average insert size 1.8 kb. Library constructed by Life		
	Technologies."		
BASE COUNT	157 a	142 c	111 g 122 t 1 others
ORIGIN			
Query Match	10.7%;	Score 343.6;	DB 11; Length 533;
Best Local Similarity	86.7%;	Pred. No. 2.9e-66;	
Matches 436;	Conservative 0;	Mismatches 60;	Indels 7; Gaps 5;
QY	1598	gaagatatttgaagaagagaaactcttgatgtgaagaagactttcaaaaatcctcaaacctgcggt	1657
Db	1	GAAAGATATTTGAACAAGAAAGAACTTGATTGAGCAAGACATTTTAAATCAATCAAAATCGCGGT	60
QY	1658	cgcacagagcttcacacatctatgagagacataacgagagcgctccctcgaatgaagattac	1717
Db	61	CG-ACAGGCTTACCAATCT-TGAGACGACAGAGGAGCGCTTTCTTAAGTCAAGGATATAC	118
QY	1718	ggctctcaagagacgcctagatgtaaaaaatcccgtagtlltcaagacacagcagatgac	1777

[illegible]

## ORIGIN

Query Match	10.78;	Score 343;	DB 11;	length 517;
Best Local Similarity	90.68;	Pred. No. 3.9e-66;		
Matches 481; Conservative	0;	Mismatches 30;	Indels 20;	Gaps 10;

OY	521	aaccccaaatgtgaattcagtttagctgtaacctgcagaataatcacccgtcatctccaagt	580
Db	1	AACCAAAATGTAATTCTA-TTAGCTACTCTGACATATATCACCGCTTATCTCCAAAG	59
OY	561	abggaataatcacatgggttcctctcttgaacaacaaagtaaacatgaatgatgccacc	640
Db	60	ATGGATTAA-TCACATATGGTTCTTCTGTGAACAACAAGAATAATGAATGATCCCCACC	118
OY	641	acagaagacatatcaatccctcagtgagctcccactgaggcgacttatattgatgacct	700
Db	119	ACAGAGAACAAATCAATCATCAG-GGCCTTCCCACCTGGACCGCTTAATGAGAGCCAGCA-	176
OY	701	gaagcctaacaagacagatggtccccagcaatccttgcacaagaatgatccccctgtgcaa	760
Db	177	----CCCTTAACAGCAGCAG-GTCCACGACATCTTGCCA---AGATGATCCCTGTGCAGA	228
OY	761	laatcgttatgttgtttaagctgtgctataataacaagitttgcctgtgtttaagagg	820
Db	229	TAAATTCGTATGTGTTAAGCTGC-----ATAATPACAAGTTTTGCCCTGTGTTAGAAGG	283
OY	821	tattactacaactcttctacatgtaagaagaagatbtatccctgagaaagattcagt	880
Db	284	TATTTCTACAACTCTCTACATGATGAAGAAAGATATTCCTCCGTG-GAAATTTTCACT	342
OY	881	gacagatacaagaacatttgcaccaagaagaacaattccatgagcctatcaagaattgca	940
Db	343	GACATATCAGAAAATTTGACCCAGAGAGAAAATTTCCATGAGCCTATCAAGACTTGCA	402
OY	941	tagtgaattactagctgtgttaaagaatgatattgycacatcgtttatgacagaactgt	1000
Db	403	TAGTGAATTTACTACTCTGTTTAAAGANTGATTTGGCACATCTGTTTATGAGCAGACTGT	462
OY	1001	aattctctgtaagagacatctctgttaccagaagtttcgaaatcgagct	1051
Db	463	AATTTCTACTGTAA-GCACATCTCTGTCTACCAAGA-TCGAAATGATGTGCT	511
RESULT	5		
LOCUS	A1791951	536 bp	mRNA EST 13-DEC-1999
DEFINITION	nm99g12.y5 NCI_CGAP_C09 Homo sapiens cdna clone IMAGE:1076422 5',		
ACCESSION	A1791951		mRNA sequence.
VERSION	A1791951.1	GI:5339667	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
COMMENT	Tumor Gene Index Unpublished (1997) Other_ESTS: nm99g12.x5 Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D., , Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLND at: www-bio.litl.gov/bdrp/image/image.html		

This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: see original entry for original citation  
information  
This 5' resequenced clone has no previous 5' data to verify this  
new read against  
Insert Length: 581      Std Error: 0.00  
Seq Primer: -40RP from Glenco  
High quality sequence stop: 453.  
Location/Qualifiers

	Source	L . . 536 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1076422" /clone_lib="NCI_CGAP_CO9" /lisse_type="colon tumor RER+" /_lib_host="DH10B" /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER+ colon tumor, and was then primed with a Not I - Oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4 ) . "			
BASE COUNT	149 a	118 c	131 g	136 t	2 others
ORIGIN					
Query Match	10.1%	Score 325:	DB 10:	Length 536:	
Best Local Similarity	89.3%	Pred. No. 4.1e-62:			
Matches 499;	Conservative	0;	Mismatches 37;	Indels 23;	Gaps 13;
OY 2627	aaaaacatgaccctgttcagaagaaagaaagaaagaaactcgttgcgggaagatcaatta 2686				
Dd 1	AAAACATGACCCTGTTCAGAGAAGAAAGAACGAAGCAACTGGCGTGAGAGATCAATT 60				
OY 2687	gagaggagcacctcgtgatccaacttcgttccttagttgccctccctccatcgacgaag 2746				
Dd 61	GAGAGAGGACCTCTGGATCCACCTTC-TTCCCTTAGGTCCTCCTCCAT-CAGCAAAAG 118				
OY 2747	agaccttcctaagcatagccctcccgaagactcgtcgggaagatttaaaaacaaa 2806				
Dd 119	AGCACTTCTCTTA-TCATGCCCTCCGAAAGACTGGCTGGAGAAAGTTT-AAAAACAA 176				
OY 2807	aatccaagataaagaccttaggttcagtcttgaaatctgaaacaactcgtcttg 2866				
Dd 177	AATCCAGGAGT-AAAGGCTTAGG---TCAGTTTAAATTGGAGACAACAT-GTCTGGC 230				
OY 2867	aaagggtgccaagagcgagcttgtctgcagagagtcaccagcgctcagacctcgggtgt 2926				
Dd 231	AAAGGGTGC---GAGAGGAGACTTGTGCTCAGGAGT-CCAGCCGTCAGCCTCGGGGTGT 286				
OY 2927	aaggtctctgaggtgtgcataaggggggctcagctcttccttgtagcccgaggtcagctg 2986				
Dd 287	AGGTTTTGAGAGTGTGCCATTGGGGCCATCACCTTCTGTGTGA-CAGAGGCTCAGCTG 344				
OY 2987	tgggccaaccaaacat 3046				
Dd 345	TGGCCACCAACACACACACACACACACACACACACACACACAAATGGGGGCAACACATC 404				
OY 3047	ccaagtaaccaagctcttacacaatatgttatagtcgtcccttcttatttctaataagcct 3106				
Dd 405	CAGTACAAGC-----TTTACAATGTTATTAGTGCTCTTTTATTTCNAAT-GCCTT 457				
OY 3107	gtccctcttaaagtattattatgttatattatcattgttctctgcgtttaatgtgat 3166				
Dd 458	GTCCCTTANAGTATTATTATTTGTATTTATTTATTTGCTTGACTCTTAATTGTGAT 517				
OY 3167	ggtaatgcataaagtgcc 3185				
Dd 518	GGTAATGCATAANAAGTGCC 536				



[illegible]

D	342	-	ATTGTGGGACCATCGCTGGCATTGTCATTTCCAGCATGATATTPCATTTGATTGCCA	399
Q	1574	ctagaagaatcaataataacaaagcgaagcatattgaaagacgaacttgattgcagag	1633	
D	400	C-AACAAGATCAATAATACAAAA-CGAACCAATTTGAMAGA-GAGAACTTGATTGCAGAG	456	
Q	1634	aacttcaaacttaaacattcgctgcgcgcgaagcttcaccatctatggagataacgag	1693	
D	457	ACTTTCAAAATCTAAACTGGCGGTG-ACAGGCTTCACCACATCT-TGAGCAGCAGGAG	514	
Q	1694	cgtctccctcaggctcaagattacgacctcca	1725	
D	515	CGCTTTCTCTAAGTCAGCATTAACGGCCTCCA	546	
R	RESULT	7		
L	LOCUS	A1949096	461 bp	mRNA
D	DEFINITION	wg19d08.xl NCI CGAP_Gas4 Homo sapiens CDNA clone IMAGE:247127 3'	EST	08-MAR-2000
A	ACCESSION	A1949096		
V	VERSION	A1949096		
K	KEYWORDS	PREDICTOR; mRNA sequence.		
S	SOURCE	A1949096.1 GI:5741406		
O	ORGANISM	Homo sapiens		
R	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
A	AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.		
T	TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/hncicgp. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LNM at: www.bio.liml.gov/bdrp/Image/Image.html Insert Length: 2926 Std Error: 0.00 Seq primer: -40UP from GIDCO High quality sequence stop: 441.		
F	FEATURES	Location/Qualifiers		
S	SOURCE	1..461		
		/organism="Homo sapiens"		
		/db_xref="taxon:9606"		
		/clone="IMAGE:247127"		
		/clone_lib="NCI CGAP_Gas4"		
		/tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"		
		/lab_host="DH10B"		
		/note="Organ: stomach; Vector: PCMV-SPOrt6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"		
B	BASE COUNT	148 a	86 c	97 g
O	ORIGIN	130 t		
Q	Query Match	9.4%; Score 301.8; DB 10; Length 461;		
B	Best Local Similarity	98.1%; Pred. No. 6.2e-57;		
M	Matches 358; Conservative	0; Mismatches 2; Indels 5; Gaps 5;		
Q	829	caactctctcatcgttaaagaagaagatcccttgagaagattcaagtgcacgat	888	
D	22	CAACGCTTCTCAATTAAGAAAGAAAGATATTCCTG-GAAGATTTCAGTGACAGTAT	80	
Y	889	cagaacaattgaccocgaagaagaacattccatgycctatcaagaacttgatagtaa	948	

```

Db      81  CAGAAACATTTCACCCAGAGAAACATTTCATGCGCTATCAAGACTTCATATGAA 140
Oy      949  ttacagactgtttaaagatgatttggcacatctgttatgacagactgtatctta 1008
Db      141  TTACTAGCTGTTTAAAGATGATTTGGCAGCATCTGTTATGACAGACTGTAATCTTA 200
Oy      1009  ctgaagcacatctctgtcaccaagatctgaatgctgcttgatgaagaatttgt 1068
Db      201  CCGTAA-CCACATCTCTGTCCACCAAGATTCGAAATGCGTGTGATGACAAAGTTTGT 259
Oy      1069  aatgtaacatgtaacaatttttggcagaaccacagaagtgacaatgaagaactgt 1128
Db      260  AATGTAACATATGTAACATTTTGGCAGAACCAAGTGACATGAGAAGACTGTGACT 319
Oy      1129  ggaagaataataaagcaattataagtaagtcgaagcaacttcttaactatgtg 1188
Db      320  -GAGAAATTAATTAAGCAA-TTAGAGTAGCTCAAGCAAC-TTCTAAACTATGATTTG 376
Oy      1189  accct 1193
Db      377  ACCCT 381

```

```

RESULT 8
LOCUS   BE870652      807 bp      mRNA      EST      20-OCT-2000
DEFINITION 60144779F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3851730 5',
            mRNA sequence.
ACCESSION BE870652
VERSION   BE870652.1 GI:10319428
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 807)
            NIH-MGC http://mgi.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-tr@mail.nih.gov
COMMENT   Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LHM9572 row: n column: 19
            High quality sequence stop: 652.
            Location/Qualifiers
                1..807

```

```

FEATURES
    source
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:3851730"
        /clone_id="NIH_MGC_65"
        /tissue_type="adenocarcinoma"
        /lab_host="DH10B (phage-resistant)"
        /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
        Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
        Average insert size 1.8 Kb. Library constructed by Life
        Technologies."
BASE COUNT  234 a      193 c      214 g      166 t
ORIGIN

```

```

Query Match      9.3%; Score 298.8; DB 11; Length 807;
Best Local Similarity 79.9%; Pred. No. 2.8e-56;
Matches 638; Conservative 0; Mismatches 102; Indels 58; Gaps 22;
Oy      2261  cgcatacactgagggtccgaggaggatgaagaaggtacccacacattcattcaagg 2320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||

```

```

Db      57  CGCATCACTGTGAGGTCGAGGG--ATGAGAGGGATACCCACCCTTTCAAGGCT 114
Oy      2321  tcacaagctacactctcgttagacaagtcagaataggagacactgtctctatccctcaat 2380
Db      115  CACAAGCTCAGCTC---TGACAAAGTCAGATATAGGAGCA--CTGGCTTCTATCCCTCAAT 169
Oy      2381  ggaaggatcttggccaaccccccttlttgaaaacagagcccccagaagcttggcaac 2440
Db      170  GGA-GGATTTCTTGCAACCTTTG-----AACAGCCAGAGCTT-CCAAC 211
Oy      2441  ctgactcaaacccaagaagatgaaaggagacatatctttagcttlttggaggagcg 2500
Db      212  CTAGCCCTC-ACCCAGAAAGACTGGAAG-GAGACATATCTC-TAGCTTTTTCAGGAGCG 268
Oy      2501  tgccttgggaatccagaagcttcttgatgtaattagaagagcctggaatataatgt 2560
Db      269  TGCC-TGGGATTCAGAGAAC-TTTTGATGCTATTTAGAAGGCTTGACATAAAATGTC 326
Oy      2561  ccatctatgggtttaaatacagttttagaatgctagagaagcaagagggccaga 2620
Db      327  AC---TATGGGGTGCACCTACAGTTTGGAA-ATGCTAGCAGCAGAGAGGGGCGAGAGA 381
Oy      2621  gagtaaaaaaacatgacctgtgtagaaaggaaagaaagaaactgtgtggaggat 2680
Db      382  G---TAAAAAACATGACCTGTGTAGAAAGAGAGAGCAAGCAAGCAACTGGGTGGAGAT 438
Oy      2681  caatlagagaagagacactgtgagatcacctgtctctctatggtccctctcatgag 2740
Db      439  CAATTAGAGAGAGGACACCTGGGATCCACCTTC-TTCCTTAGGTCCCTCTCCAT-CAG 496
Oy      2741  caaaggacacttcttaagtaatgctccctccgaagactgtgtggagaaggtttaa 2800
Db      497  CAAGAGAGCACTTCTTAA-TCATGCCCTCCGAAACACTGCGGAGAGAGGTTTAAACA 555
Oy      2801  acaaaaaatccaaagatgaagaagcttagagtcagtttgaagaattggagaacactgt 2860
Db      556  ACAAAAATTCAGAGAGT-AAGAGCTTTAGG---TCAGTTGAAATCGGAGACAAAC---T 608
Oy      2861  ctggaagaaggtgtgcaagaagcgaagctgtgtgctcaagaagatccagcgctcag 2920
Db      609  GTCGTGCAAAAGGTGCGAGAGCGGAGCTGTGCTCAGAGAT-CCAGCCGCTGACGCTCG 667
Oy      2921  ggggtgaaggtctgtgaaggtgtgcatgggggagcctagccttctgtgtgacccag 2980
Db      668  GGGTGTAAAGTCTCTAGAGGTGTCATGGGGGCTCAGCTTCTGTGATCACCCGAGGTC 727
Oy      2981  cagctgtggccacacacacacacacacacacacacacacacacacacacacacac 3040
Db      728  AGTTGTGGCACAAAAAACA-----ACACAAAACACACACACACATGGGGCACAC 778
Oy      3041  ccacatccacgttaacca 3058
Db      779  CTCATATCAAGCTTACAA 796

```

```

RESULT 9
LOCUS   AA315623      468 bp      mRNA      EST      19-APR-1997
DEFINITION ESR187445 Colon carcinoma (HCC) cell line II Homo sapiens cDNA 5',
            end similar to similar to cell surface antigen 114/AlO, mRNA
            sequence.
ACCESSION AA315623
VERSION   AA315623
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 468)
            Adams,M.D., Kerlavag,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            C.J., Lee,N.H., Kirkness,E.F., Meisnstock,K.G., Gocayne,J.D., White
            O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Whi,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald

```



OY 2812 agagagtaagagccttaggttagtcttgaanaattgagagacaacattgtcttgagcaag 2871  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 369 AGGAGT-AGGAGCCTTAGC---TCAGTTGAATGTGAGCAAACT--GTCGTGGCAAGG 316  
OY 2872 gggcagaagcgagagcttctgctcagagagctccaccgctccgctcggtgtgaagt 2931  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 315 GTGC---GAGAGGAGCCTTGTGTCTGAGAGT--CCAGCCGTCAGCCTCGGGGTGT--AGGT 261  
OY 2932 cctctgaggtgtccatcgggggctcagcctctctgtgtgaccggagctagctgtgccc 2991  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 260 TTCTGAGGTGTGCTCATTTGGGGCTTCAGCCTTCTCTGTGA-CAGAGGCTAGCTGTGGCC 202  
OY 2992 accaacaac 3051  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 201 ACCAACAACACACACACACACACACACACACACACACACAAATGGGGGACACACATCCAGTA 142  
OY 3052 taaccaagcttaacaacaattgtatagtgctccctttatttcttaataagccctgtcct 3111  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 141 CAGC-----TTTAAACAATGTATTAGTGTCTTTTATTTCTTAAT--GCCCTGTCTT 89  
OY 3112 cttaaaagtatttatttcttatttatttcttctgactgttaattgtgaattgtaa 3171  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 88 CTTAAAGTATTATTATTGTTATTTATTTATTTCTTCTGACTGTATTTGTGAATGTGA 29  
OY 3172 tgcataaagtgcttctgttagatg 3197  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 28 TCAATAAAGTGCCTTTGTAGATGC 3

## RESULT 11

AW751697 477 bp mRNA EST 28-APR-2000  
LOCUS IL4-CT0081-130999-001-C05 CT0081 Homo sapiens cDNA, mRNA sequence.  
DEFINITION AW751697  
ACCESSION AW751697  
VERSION AW751697.1 GI:7666629  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 477)  
Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R.,  
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
Brunstein, A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

CONTACT: Simpson A.J.G.  
20202663  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/ILICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL4&t2=IL4-CT0081-  
130999-001-C05&t3=1999-09-13&t4=1)  
Seq primer: puc 18 forward.

## FEATURES

## SOURCE

1..477  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CT0081"  
/dev\_stage="Adult"  
/note="Organ: colon; Vector: pUC18; Site\_1: SmaI; Site\_2:  
SmaI; A mini-library was made by cloning products derived

from ORFESTS PCR (U.S. Letters Patent application No. 196  
,716 - Ludwig Institute for Cancer Research) profiles  
into the pUC 18 vector. Reverse transcription of tissue  
mRNA and cDNA amplification were performed under low  
stringency conditions."

BASE COUNT 139 a 107 c 114 g 117 t  
ORIGIN

Query Match 8.1%; Score 260.6; DB 10; Length 477;

Best local similarity 86.2%; Pred. No. 9,7e-48;

Matches: 432; Conservative 0; Mismatches 44; Indels 25; Gaps 12;

OY 1068 taatgaacaatgtaacaatttggcagaacacacaaagtgaacaatgagaagactgtgac 1127  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 2 TGATGTAAACAAATAGTGACAATTTGGCAGAAACACAAAGACAAATGAGAAGACTGGCAG 61

OY 1128 tggagaanaattaaagaacattataagtagctcaagaacatttctaactatgattg 1187  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 62 T-GAGAAATCAATTAAGCAATT-AGAAAGTGCTCAAGCAAC-TTCTTAACATATGATTT 118

OY 1188 gaccctgtcgtgtgattgattgaaggtcgggaaccaagactggtgtgactgctc 1247  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 119 GACCC-----TTGGAGTGTAATATTATGCTGTAAACCAACTG--CGATGACTGCTC 170

OY 1248 aatgggttagcattgcgaatgtgaacatgctgacctgacaagacctaaacacagagcct 1307  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 171 AAT-GGTTTAGCATGCACTG---CAAAATCTGACCTGCAAGAGGCTTAACCCACAGAGCCCT 226

OY 1308 ttctgtgtgcttcacagctcagaagtgcttgatgctctgtaagaacagagaagaagat 1367  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 227 TTCTGCTTGTCTTCACATCTCA-AGTGTCTGATGCTTGAACGACAGCAAGC-AAT 284

OY 1368 gcttaataagaagatggtgtgggtccctcgacgtgtgtgctgtgctgtctacca 1427  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 285 GCTTAATTAAGAAAGAGTGTGGGGCCCTGAGTGTGCTCGTG-----CCGGCTACCA 339

OY 1428 ggaagatgctaattggaactgccaagaagtgtgcatttggctacagtgaactgactgta 1487  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 340 GGAAGATGCTAATGGAAGACGCCAAAGTGTGCATTT-GGCTACAGAGGACTGACGTGA 398

OY 1488 aggaacaatttcgctgactcactcattttgtggagacatcgtcgtgacttcattct 1547  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 399 AGGACAAATTCAGCTGATCTCTACT--ATTGTGGGACCACTGCTGGCATTTGCTTCT 456

OY 1548 cagcatgataatgcatgatt 1568  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 457 CAGCATGATTAATTCATTTGAT 477

## RESULT 12

AI732655 454 bp mRNA EST 13-DEC-1999  
LOCUS nm99412.x5 NCI\_CGAP\_C09 Homo sapiens cDNA clone IMAGE:1076422 3',  
DEFINITION AI732655  
ACCESSION AI732655  
VERSION AI732655.1 GI:5053768  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 454)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index

Unpublished (1997)  
CONTACT: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.

## JOURNAL

## COMMENT

cdna Library Preparation: M. Bento Soares, Ph.D.  
cdna Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)

This read is a RESEQUENCE of a previously sequenced human clone  
 Original clone citation: National Cancer Institute, Cancer Genome  
 Anatomy Project (CGAP), Tumor Gene Index  
 This read has been verified (found to hit its original self in the  
 correct orientation)  
 Insert Length: 581 Std Error: 0.00  
 Seq primer: -400P from Gibco.

## FEATURES

## source

```
1..454
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Cg9"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo (Soares4
)."
location/Qualifiers
```

BASE COUNT 125 a 102 c 104 g 123 t  
 ORIGIN

Query Match 7.9%; Score 252.8; DB 10; Length 454;  
 Best Local Similarity 87.6%; Pred. No. 5.3e-46;  
 Matches 417; Conservative 0; Mismatches 37; Indels 22; Gaps 12;

```
QY 2717 ccttaagtccttcctccatgcagaaagagcactctcctcaagtcacccctccgaag 2776
|||
DB 454 CCTTAGTCCCTCCTCCAT-CAGCAAAAGAGCACTTCTTAA-TCATGCCCTCCCGAAG 397
|||
QY 2777 actgctcggagaaggtttaaaaaaataccagagtaagagccttagggctcagt 2836
|||
DB 396 ACTGCTGGGAGAGGTTT-AAAAACAAAATCCAGAGT-AAAGCCTTAGG--TCA 342
|||
QY 2837 ttgaaatctggagaaactctctgcaagagtgccagcgagcttgctc 2896
|||
DB 341 GTTTGAATGGAGACAAACT-GTCTGGCAAGGCTGC--GAGAGGAGACTTGTGCTC 287
|||
QY 2897 agagtcacagcgctcagcctcgggtgtaagctctcctgagtggtgcagcgagcctc 2956
|||
DB 286 AGGAGT-CCAGCCCTCCAGCTCGGGGTG-AGGTTTGTAGGTGTGCTCATTTGGGGCTC 229
|||
QY 2957 agcctctctgtgaccgagctcagctgtgcaaccaacaacaacaacaaca 3016
|||
DB 228 AGCCTTCTGGTGA-CAGAGGCTCAGCTGTGGCCACCAACACACACACACACA 170
|||
QY 3017 accacaacaacaatgggggcaaccacacatccacgtaaccaagctttaacaatgtta 3076
|||
DB 169 ACACACACACAAATGGGGGCAACCATCCAGTACAGCAAGC-----TTTTCAAATGTTA 116
|||
QY 3077 ttatgtcccttttatcttaataagccctgtctcttaaaagttaattttgtatt 3136
|||
DB 115 TTATGTCTCTTTTATTTCTTAAT-GCCTTGTCTCTTAAAGTTATTTTATTTGTATTT 57
|||
QY 3137 ataattgtctctgactgtaattgtgtaattgtaattgcaataaagtccttgta 3192
|||
DB 56 ATTAATTGTCTTGTAGCTGTAAATGTGAATGTAATGAATGAATGAATGAATGAATGA 1
```

RESULT 13  
 A1573096/c  
 LOCUS A1573096 449 bp mRNA EST 14-MAY-1999

DEFINITION trf9e02.x1 NCI\_CGAP\_Pan1 Homo sapiens cDNA clone IMAGE:2223578 3',  
 mRNA sequence.  
 ACCESSION A1573096  
 VERSION A1573096.1 GI:4536470  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 449)

AUTHORS

TITLE

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgapb-remail.nih.gov  
 Life Technologies catalog #: 11548-013  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
 Insert Length: 2580 Std Error: 0.00  
 Seq primer: -400P from Gibco  
 High quality sequence stop: 410  
 POLYA-No.

FEATURES  
 source location/Qualifiers

```
1..449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Pan1"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.72 kb. Life Technologies catalog #:
11548-013"
location/Qualifiers
```

BASE COUNT 125 a 104 c 99 g 121 t  
 ORIGIN

Query Match 7.8%; Score 250.4; DB 10; Length 449;  
 Best Local Similarity 87.6%; Pred. No. 1.8e-45;  
 Matches 403; Conservative 0; Mismatches 36; Indels 21; Gaps 11;

```
QY 2738 cagcaagagcactctcttaagtcacgctcccgaaagctgctggagaagttta 2797
|||
DB 441 CAGCAAGAGAGACTTCTTAA-TCATGCCCTCCGAAAGACTGCGGAGAGGTTT-A 384
|||
QY 2798 aaaaacaaaacccagagtaagagccttaggcagctttgaaatctggagaaact 2857
|||
DB 383 AAAACAAAATCCAGGAGT-AAAGCCTTAGG--TCAGTTGAATTTGGAGACAAACT 328
|||
QY 2858 tctctgcaagaggtgcagagcgagctgtgtctcagaggtccagccgtccagcc 2917
|||
DB 327 --GTGTGGCAAAAGGTGC--GAGAGGAGCTGTGTCTCAGAGAGT-CCAGCGCCCAAGC 274
|||
QY 2918 tcgggtgtaagctctcgtgagtgcaatggggcctcagcctctctgtgacccgag 2977
|||
DB 273 TCGGGGTGT-AGGTTTGTAGAGTGTGCTATGGGCGCTCAGGCTTCTGTGTGA-CAGAG 216
|||
QY 2978 gctcagctgtgcaacaacaacaacaacaacaacaacaacaacaacaataatggggc 3037
|||
DB 215 GCTCAGCTGTGGCCACCAACACACACACACACACACACACACACACCAATGGGGC 156
|||
QY 3038 aaccacatccaagtaaccaagcttaacaacaatgtattagtgctcctttattct 3097
|||
DB 155 AACCAATCCAGTACAGC-----TTTACAAATGTTATTTAGTCTCTTTTATTTCT 102
|||
QY 3098 aatagccctgtcctcttaaaagtatttattgttattattattgtcttgactgta 3157
|||
DB 101 AAT-GCCTTGTCTCTTAAAGTATTATTTATTTATTTATTTATTTATTTATTTATTTATTT 43
```

```

QY 3158 attggaatgtaatgcaataaagtcctgtttagatgg 3197
Db 42 ATTGTGAATGTAATGCAATAAAGTCCTTTGTAGATGG 3

RESULT 14
AM363038/c 531 bp mRNA EST 04-FEB-2000
LOCUS RC0-CT0301-271199-031-G12 CT0301 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM363038
ACCESSION AM363038
VERSION AM363038.1 GI:6867688
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 531)
AUTHORS HCGP http://www.ludwig.org.br/ORESTES.
TITLE The FAPESP/LICR Human Cancer Genome Project
JOURNAL Unpublished (1999)
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL:
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC0&c2=RC0-CT0301-
271199-031-G12&t3=1999-11-27&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 14
High quality sequence stop: 399.

FEATURES
source
location/Qualifiers
1..531
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0301"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site.1: SmaI; Site.2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 112 a 155 c 117 g 147 t
ORIGIN
Query Match 7.7%; Score 246.6; DB 10; Length 531;
Best Local Similarity 87.0%; Pred. No. 1.3e-44;
Matches 481; Conservative 0; Mismatches 44; Indels 26; Gaps 18;

QY 2450 acccaagaagactggaaagagacatacttccagcttttccagaagcgctgccttggg 2509
Db 525 ACCCAAGAAGACTGGAAA-GAGACATATC-TCTCAGCTTTTTCAGAGAGCGGTGCC-TGGG 469

QY 2510 aatccagaagcttttgcgttaataagaagcctcgagctataataatgcatctatg 2569
Db 468 AATCCAGGAGAC-TTTTGTGAGTCTAATTAGAGAGCCCTGAGCTAAATGTCAC----TAT 414

QY 2570 gggcttaatcacagctttgaacatgctagagcgagaacgggcccagaagataaaa 2629
Db 413 GGGGTGCACTCTACAGCTTTTGAA-ATGCTAGAGAGCAGAGAGGCGCAGAGAG----TAAA 358

QY 2630 aacatgacctggttagaaggaagagagcaaaagaaactggtggtggagagatcaatagag 2689
Db 357 AACATGACCTGTGTAGAGAGAGAGAGCAAGCAAACTGGGTGGGAGAGATCATATTAGAG 298

QY 2690 aggaagacctggtgataccaccttcgttcttagtgcctcctccatgcagcaagagagc 2749

```

```

Db 297 AGAGAGCACTGGGATCCACCTTC-TTCCTTAGGTCCCTCCCTCCAT-CAGCAAGAGAGC 240
QY 2750 actctctaaatcatgcccccacgaagactgctgggagagaggtttaaataacaaaaaat 2809
Db 239 ACTTCTCTAA-TCATGCCCTCCCGAAGACTGCTGGAGAGAGTTT-AAATAACAAAAAT 162
QY 2810 ccagaagtaaaagagccttaggtcagtttgaataattggagacaactgttctggcaaa 2869
Db 181 CCAGAGAGT-AAGAGCCTTAGC---TCAGTTTGAATTGGAGACAACCT-GTCTGGCAAA 128
QY 2870 ggggtccaaagcggagcgtgtgtcaggaagatcccaagccgtccagcctggggtgaag 2929
Db 127 GGGGTGC--GAGAGGAGAGCTTGTGCTCAGAGAT-CCAGCCGCCACGCTCGGGGTGT-AG 73
QY 2930 gctctgaagctgtgcacatggggcgcccaagcctctcgtgtgaccgaggtcagctgtg 2989
Db 72 GTTCTGAGGTGTGCACTTGGGGCTCAGCCTTCTGTGTGA-CAGAGGCTCAGCTGTGG 14
QY 2990 ccaccaacacaca 3002
Db 13 CCGCAACACTCA 1

RESULT 15
AA581944 431 bp mRNA EST 26-SEP-1997
LOCUS AA581944/c
DEFINITION n40b07.s1 NCI_CGAP_GC5 Homo sapiens cDNA clone IMAGE:1086325 3',
mRNA sequence.
ACCESSION AA581944
VERSION AA581944.1 GI:2360622
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 431)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 1176 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 430.

FEATURES
source
location/Qualifiers
1..431
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1086325"
/clone_lib="NCI_CGAP_GC5"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site.1: EcoRI; Site.2: XhoI
; Cloned unidirectionally. Primer: Oligo dT. Mixed germ
cell tumors. 5' adaptor sequence: 5' GAATTCGGACGACG 3' 3'
adaptor sequence: 5' CTCAGTGTGTTTTTTTTTTTTTTT 3' Average
insert size: 0.7 kb."

BASE COUNT 122 a 102 c 92 g 115 t
ORIGIN
Query Match 7.6%; Score 243.6; DB 10; Length 431;
Best Local Similarity 87.8%; Pred. No. 6e-44;

```



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:07:20 ; Search time 519.52 Seconds  
(without alignments)  
5290.623 Million cell updates/sec

Title: US-09-867-034-4

Perfect score: 3206

Sequence: 1 ttccgctcgcagctgaact.....ttgttagatggaataaaaaa 3206

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing:

Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1030	32.1	2265	AA129139	Colon tumour relat
2	1007.4	31.4	2263	AA252937	Human prostate tum
3	655.6	20.4	1668	AA277886	Human cancer assoc
4	655.6	20.4	1668	AA2500838	Human colon cancer
5	654.4	20.4	1667	AA341197	Human colon cancer
6	365	11.4	536	AA35580	Human colon cancer
7	247.8	7.7	313	AA87786	EST clone ES206
8	200.2	6.2	417	AA17847	CDNA encoding huma
9	200.2	6.2	417	AA128585	Colon tumour relat
10	187.6	5.9	401	AA17762	CDNA encoding huma
11	187.6	5.9	401	AA128500	Colon tumour relat

12	148	4.6	457	22	AAH34198	Human colon cancer
13	136.8	4.3	295	16	AA122467	Human gene signatu
14	123.2	3.8	593	21	AAA16127	Human colon cancer
15	116.8	3.6	620	21	AAA16148	Human colon cancer
16	109.4	3.4	570	21	AAA16119	Human colon cancer
17	105.8	3.3	936	22	AA158252	Oligonucleotide D1
18	105.8	3.3	936	22	AA158254	Oligonucleotide D1
19	105.8	3.3	936	22	AA158257	Oligonucleotide D1
20	105.8	3.3	936	22	AA158259	Oligonucleotide D2
21	105.8	3.3	936	22	AA158262	Oligonucleotide D2
22	105.8	3.3	936	22	AA158255	Oligonucleotide D1
23	103.6	3.2	936	22	AA158252	Oligonucleotide D1
24	103.6	3.2	936	22	AA158254	Oligonucleotide D1
25	103.6	3.2	936	22	AA158257	Oligonucleotide D1
26	103.6	3.2	936	22	AA158259	Oligonucleotide D2
27	103.6	3.2	936	22	AA158262	Oligonucleotide D2
28	103.6	3.2	936	22	AA158255	Oligonucleotide D1
29	53	1.7	5163	19	AA158255	Oligonucleotide D1
30	53	1.7	5163	21	AA158255	Oligonucleotide D1
31	53	1.7	5163	21	AA158255	Oligonucleotide D1
32	52.2	1.6	5318	19	AA158255	Oligonucleotide D1
33	50.4	1.6	5318	19	AA158255	Oligonucleotide D1
34	50.4	1.6	5318	19	AA158255	Oligonucleotide D1
35	48.4	1.5	7334	21	AA158255	Oligonucleotide D1
36	48.4	1.5	867	20	AA158255	Oligonucleotide D1
37	48.4	1.5	1397	20	AA158255	Oligonucleotide D1
38	48.4	1.5	1397	20	AA158255	Oligonucleotide D1
39	47	1.5	244	22	AA158255	Oligonucleotide D1
40	45.6	1.4	5912	21	AA158255	Oligonucleotide D1
41	45	1.4	163	20	AA158255	Oligonucleotide D1
42	44.2	1.4	244	22	AA158255	Oligonucleotide D1
43	44.2	1.4	573	22	AA158255	Oligonucleotide D1
44	44.2	1.4	573	22	AA158255	Oligonucleotide D1
45	43.6	1.4	436	22	AA112933	Probe #6015 used t

#### ALIGNMENTS

RESULT 1  
ID AA129139 standard; cDNA: 2265 BP.  
XX AA129139;  
DT 12-OCT-2001 (first entry)  
XX  
DE Colon tumour related longer determined cDNA sequence for C794P.  
XX  
KW Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
KW immunogenic; gene therapy; vaccine; colonic cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200149716-A2.  
XX  
PD 12-JUL-2001.  
XX  
XX  
XX 29-DEC-2000; 2000WO-US35596.  
XX  
XX 30-DEC-1999; 99US-0476296.  
XX 10-JAN-2000; 2000US-0480321.  
XX 15-FEB-2000; 2000US-0504629.  
XX 06-MAR-2000; 2000US-0519444.  
XX 19-MAY-2000; 2000US-0575251.  
XX 29-JUN-2000; 2000US-0609448.  
XX 28-AUG-2000; 2000US-0649811.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;  
XX King GE, Wang T, Jiang Y;  
XX



```
OY 2421 gcccacagagcttggacactagcctcaaccccaagaagactggaaaagacatcctt 2480
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1514 acagccacagagcct-gcaacctagcctc-acccaagaagactggaaa-gagacatacc-t 1569
OY 2481 ttcaagcttttccaggagcgctgaccttgggaatccagaagcgtttttgatcctaattagaa 2540
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1570 ctcaagcttttccaggagcgctgaccttgggaatccagaagcgtttttgatcctaattagaa 1627
OY 2541 ggccttgactataataatgtccatctatgggttttaactcaagttttgaaactgcta 2600
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1628 ggccttgactataataatgtccatctatgggttttaactcaagttttgaa-attgcta 1682
OY 2601 gggagcagagacggggccagagagatgaataaaacatgacctggttaagaagaagagagcaaa 2660
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1683 gggagcagagaggggagagagagagagagagagagagagagagagagagagagagagag 1739
OY 2661 ggaactgtgtgtggagagagagagagagagagagagagagagagagagagagagagagag 2720
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1740 ggaactgtgtgtggagagagagagagagagagagagagagagagagagagagagagagag 1798
OY 2721 aggtccctccctccatctgacagagagagagagagagagagagagagagagagagagag 2780
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1799 aggtccctccctccatctgacagagagagagagagagagagagagagagagagagagag 1856
OY 2781 gctgggagagaggtttaaaaaaacaataccagagagagagagagagagagagagagagag 2840
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1857 gctgggagagaggttt-aaaaaacaataccagagagagagagagagagagagagagagag 1911
OY 2841 aaaaattggagacaaactgtcttctggcaaaagggtgccaagagcgagagcttctgtctcaga 2900
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1912 gaatttgagagacaaact-gtctggcaaaagggtgc---gagagagagagcttctgtctcaga 1966
OY 2901 gtcccaagcgctccagagctcggtgtgtaaggtctctgaggtgtgccaatgggggagagc 2960
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1967 gt-ccaagcgctccagagctcggtgtgtaaggtctctgaggtgtgccaatgggggagagc 2024
OY 2961 ttctctgttaccggagagagagagagagagagagagagagagagagagagagagagagag 3020
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2025 ttctctgttaccggagagagagagagagagagagagagagagagagagagagagagagag 2083
OY 3021 cacacacaaatgggggagagagagagagagagagagagagagagagagagagagagagag 3080
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2084 cacacacaaatgggggagagagagagagagagagagagagagagagagagagagagagag 2137
OY 3081 tgtcccttttatttcaataagccctgtcctcttaaaagttaattttattttattta 3140
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2138 tgtcccttttatttcaataagccctgtcctcttaaaagttaattttattttattta 2196
OY 3141 ttgtcttctgactgtaattgtaattgtaattgtaattgtaattgtaattgtaattgtaatt 3200
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 2197 ttgtcttctgactgtaattgtaattgtaattgtaattgtaattgtaattgtaattgtaatt 2256
OY 3201 aaaaaa 3206
    |||||||
Db 2257 aaaaaa 2262
```

## RESULT 2

AA252937 standard; cDNA; 2263 BP.

AA252937,

14-MAR-2000 (first entry)

Human prostate tumor cDNA library derived EST fragment #80.

Pancreas; tumor; EST; expressed sequence tag; human; cytosolic;

treatment; ds.

Homo sapiens.

XX

```
PN DE19820190-A1.
XX
PD 04-NOV-1999.
XX
PF 28-APR-1998; 98DE-1020190.
XX
PR 28-APR-1998; 98DE-1020190.
XX
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX
PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;
XX WPI: 1999-621386/54.
DR P-PSDB: AAY74052, AAY74053, AAY74054.
XX
PT New human nucleic acid sequences from pancreatic tumors, and related
PT proteins -
XX
PS Claim 2; Page 247-248; 502pp; German.
XX
CC This invention describes novel polypeptides and their encoding nucleic
CC acid derived from human pancreatic tumor tissue which have cytosolic
CC activity. The sequences are also useful in producing pharmaceutical
CC compositions for treatment of pancreatic tumors. AA252858-253014
CC represent expressed sequence tag (EST) fragments derived from a human
CC pancreatic tumor cDNA library and which encode the proteins represented
CC in AAY73814-Y74252.
XX
SQ Sequence 2263 BP; 689 A; 484 C; 520 G; 570 T; 0 other;

Query Match 31.4%; Score 1007.4; DB 20; Length 2263;
Best Local Similarity 83.0%; Pred. NO. 1.3e-269;
Matches 1979; Conservative 0; Mismatches 261; Indels 145; Gaps 67;

OY 822 attactaactctctcatgtaagaagaagagatctccctggagaagattcagtg 881
    ||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 attacgacaactctctcatgtaagaagaagagatctccctggagaagattcagtg 59
OY 882 acagatcagaacaatttgaccagagaagaagaacattccatgagctatcaagacttgat 941
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 60 acagatcagaacaatttgaccagagaagaagaacattccatgagctatcaagacttgat 119
OY 942 agtaaatctactgctgtttaagaatgatttggacatctgtttatggagacagtta 1001
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 120 agtaaatctactgctgtttaagaatgatttggacatctgtttatggagacagtta 179
OY 1002 attctactgtaagggacatctctgcacacagaattctgaatcgctgtgatgacaag 1061
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 180 attctactgtaagggacatctctgcacacagaattctgaatcgctgtgatgacaag 236
OY 1062 ttctgttaatgtaacaatagtaacaatttggcagaacaacaaagtgtacaatgagaag 1121
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 237 -ttgtttaatgtaacaatagtaacaatttggcagaacaacaaagtgtacaatgagaag 295
OY 1122 ttgtactggagaataatgaataaagaatttaagttaagcttaagcaactttctaaacta 1181
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 296 ttgtactggagaataatgaataaagaatttaagttaagcttaagcaactttctaaacta 352
OY 1182 tgattggacccctgtcggtgtgattgtagggcgtgggagacaaagactgtgtagtac 1241
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 353 tgattggacccctc-----gggtgtgattatattatggtcgttaacagactgc--ggaatgac 404
OY 1242 tgcctcaatgggttagcagatgtagcaaatgtcgtcagctgtcaaaaggcctaaccacag 1301
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 405 tgcctcaat-ggtttaagatgcatgcatgc--aaatctgacgtgcaaaaggcctaaccacag 461
OY 1302 agcccttctgctgtgctcagagctcagagtgcttcgtatgcttcgcaacggagacacaa 1361
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 462 agcccttctgctgtgctcagagctcagagtgcttcgtatgcttcgcaacggagacacaa 520
OY 1362 gcaatgcttaataaagaagtggtgtgggtccctcgtcagtgctgtgctgctccggt 1421
    ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
```

Db	521	gc aatgcttaataaagaagatggtgtgtggcccttgatgtgtcgtgtg-----cccg	574
Qy	1422	ctcaccaaggaaagatgcttaattggaaactgcacaaagtgatgatttgggtctaacgttgatctg	1481
Db	575	cttaccgaagaagatgcttaattggaaactgcacaaagtgatgattt-ggtacagtgatctg	633
Qy	1482	actgtaagaaacaaatttcagctgtgactctcaactatttgggtgacacatctgcgtgcatgtt	1541
Db	634	actgtaagaaacaaatttcagctgtgactctcaact--atttgggtgacacatctgcgtgcatgtt	691
Qy	1542	catcttcagctgtgataatttgattgattgtgtcactatgacaaagatcaataataacaaagcgag	1601
Db	692	catcttcagctgtgataatttgattgattgtgtcact-aggcaagatcaataataacaa--cgag	749
Qy	1602	catatttgaagaacgagaaacttgatgtatgtacgaaagacttccaaactctaaacttcgtctgc	1661
Db	750	catatttgaaga--gagaacttgtatgtacgaaagacttccaaactctaaacttcgtctgc-a	807
Qy	1662	caggtcttcaacatctatgtatgagcataaagagcgtcttcctcaggttcagatgtacggtc	1721
Db	808	caggtcttcaacatctt-tggagcgaaagagagcgtcttcctcctaaggtctcagataaagcgc	866
Qy	1722	lcccaagagagcgcgttaagatcaaaaatcccgtagtttcaagaaacagacgatactgcgcg	1781
Db	867	lcca--gagacagccagatgtaaaaatccc--tattcaagaatacagccagatgt--cccg	919
Qy	1782	gcttgcatactatagaatccatcatgaatgttggaaacccgcacatgcccccaacataatgtaca	1841
Db	920	cccttgacattatgaat--cataagatgttgaaccgcgcatagtcccccaacaaigtataca	978
Qy	1842	tatctatattctctgcaggtgttttagacaagaactgtcatgtgagaagtgtgacacacgtlaag	1901
Db	979	gctatttattt-----agagtgtttagaagaagactgtgttgagaagtgtgacacca--gtaaa	1030
Qy	1902	actcgtgcgtcccggaagtgttcttcttcctccactgataatgctgcaggtcctcactgcgaat	1961
Db	1031	gactcgtgcgtcccggaagtgttcttcttcctccactgataatgctgcaggtcctcactgcgaat	1090
Qy	1962	ggcaacgctgtgcacatgtctctgcacaaagacatccacgcgtcacttgtctaaataaagaact	2021
Db	1091	gtgaatgtttgcaag-----aatccagctcacttgct--aataagaactc	1135
Qy	2022	atgacataaacaatgtatgactgcgaatgtatgtacgtgcgtgcacagagaagtgtgttctctc	2081
Db	1136	atgacatata--aatgtatgtatgactgtatgtacgtgcgtgcacagaga--ggtgtgttctctc	1190
Qy	2082	aatacgaatacaaaagtactctggacaaatcgaatcgttaggtgtgttcttctaattcttccctgc	2141
Db	1191	aatacgt--acaaagtactctggacaaatcgaatcgttaggtgtgttcttctaattcttccctgc	1247
Qy	2142	taagggaacaaagaccccatcttccaaactctgaaagacgtccccaagatctgtctgtc	2201
Db	1248	tagggcaacaaagaccc--atttccaatcttagaggaag--ctcccaagatctgtct--tg	1300
Qy	2202	ccctgggccaacaacatgctcttcttgaatgaatgtgaacttaactccctggagcagacatacc	2261
Db	1301	ctccctgggccaacaacatgctcttcttgaatgaat--tgaccttaactccctggagacata--c	1356
Qy	2262	gcatcaaacgtgtggaggtctgcgaaggg--atgagaagaggtataccacacacttcaaggtgc	1414
Db	1357	gcatcaaacgtgtggaggtctgcgaaggg--atgagaagaggtataccacacacttcaaggtgc	1414
Qy	2322	cacaagctacaactctgttgcgaagtacgaataagggacacgtctctataccctccaatg	2381
Db	1415	acaagatctacactctc--tgacaaggtcagaataagg--aacgtctctataccctccaatg	1469
Qy	2382	gaggagatcttgcgcacaaacccctcttcttggaaaacagcgccccaagatcttggcaac	2441
Db	1470	ga--gagatttctgga-----ccttggaaacagcccaagatctt-gcaacc	1511
Qy	2442	tagcttcaacccaagaagactgtgaaagagacatatctttagcttcttcaagagcgt	2501
Db	1512	tagcttcaacccaagaagactgtgaaagagacatatctttagcttcttcaagagcgt	1568

QY	2502	gctctgggaatccagaacgctttttatgcttaataagaagccctgactataataatgctc	2501
Db	1569	gcc-tgggaatccaggaac-tttttatgcttaataagaagccctgactataataatgctc	1626
QY	2562	catctatgggttttaatctacagcttttttaaatgcttagggggagagaacggcgccag	2621
Db	1627	ctat----ggggctgactctaccgctttttgaa-atgcttaggggcagaaaggggcaggag	1601
QY	2622	agtaaaaaaatatgacctgtgtagaaggaagaagcgcaaaagaaactgggtgggagagatc	2601
Db	1682	--taaaaaaatatgacctgtgtagaaggaagaagcgcaaaagaaacggggtgggagagatc	1738
QY	2682	aattagagagagagcaccctgggatccaacctgtgttccctaggttccctctcatgagc	2741
Db	1739	aattagagagagagcaccctgggatccaacctt-ttccctaggttccctctcatc-t-cagg	1786
QY	2742	aaagagagactctctctatgcatgacctcccgaaagcttgctctggagagaaggtttaaaaa	2801
Db	1797	aaagagagactctctctatgcatgacctcccgaaagcttgctctggagagaaggttt-aaaaa	1854
QY	2802	caaaaaatccagsgatlaaagagccttagggctcagtttggaaaaatgagacaaactgtc	2861
Db	1855	caaaaaatccagsgat-aagagccttagg--tcagtttggaaatgagagacaaact--gt	1908
QY	2862	ttggcaaaaggctgccaagaagcggagctgtgtgtctgagagtcgccagcgctcagcctcg	2921
Db	1909	ctggcaaaagggttgc--ggaagaggagctgtgtgtctcagaggt-cgaagcgctccaagcctcg	1964
QY	2922	ggctlaagctctctcgaagcttgccaatgggggacctcagcctctctgtgtaccggagctc	2981
Db	1965	gggtt-agggttctcgaaggtgtgcattggggacctcagcctctctctgtgtacc-gagagctc	2022
QY	2982	agctgtgccaacaac	3041
Db	2033	agctgtgccaacaac	2082
QY	3042	caatccaggaacaaagctttaacaaatgttatatgtatgtcccttttatatttataa	3101
Db	2083	acatccaggtacaagc-----tttaacaaatgttatatgtatgtcccttttatatttataa	2135
QY	3102	gccctgtcctcttaaaagtattttatctgttatattatctgtctctgactgttaattg	3161
Db	2136	gccctgtcctcttaaaagtattttatctgttatattatctgtctctgactgttaattg	2195
QY	3162	tgaatgttaatgcataaagtgccttgtgtagatggaaaaaaa	3206
Db	2196	tgaatgttaatgcataaagtgccttgtgtagatggaaaaaaa	2240
RESULT 3			
AAC77886			
ID	AAC77886 standard; cDNA; 1668 bp.		
NC	AAC77886;		
AC			
XX			
DT	08-FEB-2001 (first entry)		
XX			
DE	Human cancer associated gene sequence SEQ ID NO:260.		
XX			
KW	Human; cancer associated gene; cancer antigen; detection; cancer;		
KW	diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;		
KW	antidiabetic; antisthmatic; antihemmatic; antiarthritic; antiviral;		
KW	antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;		
KW	dermatological; neuroprotective; thrombolytic; coagulant; nootropic;		
KW	vasotropic; antiporiatic; antiangiogenic; gene therapy; inflammation;		
KW	immune reaction; haematopoietic cell disorder; autoimmune disorder;		
KW	allergic disorder; graft versus host disease; organ rejection;		
KW	haemostatic; thrombolytic; cardiovascular disorder; infection;		
KW	neurological disease; drug screening; ss.		
XX			
OS	Homo sapiens.		

PN W020005350-A1.  
XX 21-SEP-2000.  
XX 08-MAR-2000; 2000WO-US05882.  
XX 12-MAR-1999; 99US-0124270.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM.  
XX WPI; 2000-587533/55.  
DR P-PSDB; ABA3677.  
XX  
PT Novel isolated nucleic acids comprising sequences encoding peptides  
XX useful for treating or diagnosing e.g. cancer -  
PS Claim 1; Page 842; 2352pp; English.  
XX  
CC AAC77607 to AAC78448 encode the human cancer associated proteins given  
CC in ABA43398 to ABA44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytostatic; proliferative; vulnery; immunomodulator;  
CC antidiabetic; antisthmatic; antirheumatic; antiarthritic;  
CC antiinflammatory; antihypoid; antiallergic; antibacterial; antiviral;  
CC dermatological; neuroprotective; cardiac; thrombolytic; coagulant;  
CC neurotropic; vasotropic; antipruritic and antineoplastic. The  
CC polynucleotides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and ABA44240 represent sequences used in the exemplification of  
CC the present invention.  
XX  
SQ Sequence 1668 BP; 503 A; 362 C; 389 G; 413 T; 1 other;  
Query Match 20.4%; Score 655.6; DB 21; Length 1668;  
Best Local Similarity 80.7%; Pred. No. 7.5e-172;  
Matches 1426; Conservative 1; Mismatches 220; Indels 119; Gaps 52;  
QY 1441 gggagctgcacaaagtgcatttgggtacagtgactgcacttaagacaatttca 1500  
DB 1 gggagctgcacaaagtgcattt-gctcacagtgactgcactgcgttaagacaatttca 59  
QY 1501 gctgacctcactatttgttggcaccatcgctgcatgttcattctcagatgataatt 1560  
DB 60 gctgacctcact--attgttggcaccatcgctgcatgttcattctcagatgataatt 117  
QY 1561 gcatgtatgtcacttagcaagatacaatacaaaagcgaagacatattgaagacagagac 1620  
DB 118 gcatgtatgttsac-agcaagatacaatacaaaa-cgaagacatatttgaagaa-gagacac 174  
QY 1621 ttgtttgacgaagacttccaatctaaactgctgcacagcttcacacactctatg 1680  
DB 175 ttgtttgacgaagacttccaatctaaactgctgcg-acagcttcacacactct-tg 232  
QY 1681 gacataacgagagcttccctcagtcagatcagctcccaagacagccctagat 1740  
DB 233 gacataacgagagcttccctcagtcagatcagctccca--gacacagcagat 290  
QY 1741 gcaaaaatcccgtagttccaagacacagcagatgcccgcgctgactatagatcca 1800  
DB 291 gcaaaaatcc---tatccaagcacaagcagatg-cccgcgctgactatagat--ca 344

QY 1801 tcaaatgtggaaccgcgcatgcccccaacatagtacatactattctactagcagt 1860  
DB 345 taagaatgtggaaccgcgcatgcccccaacatagtacatactattt-----aga 398  
QY 1861 gtttagacaagactcagtcaggaagttagcaccacagtaaaagactctggcctccggagtt 1920  
DB 399 gttttagaagaagactcagtcaggaagttagcaccac--gtaaaagactcggcctccggggtt 456  
QY 1921 tctcttcacatcagacatactgcagtcctccatctcactcagatgcgaagtgtgcaatgtc 1980  
DB 457 tctcttcacatcagacatactgcagtcctccatctcactcagatgttgaatgttgca----- 511  
QY 1981 ttgcaaacgacatccagctcactctgctcaaaaataagaatctatcacttaacagtagct 2040  
DB 512 -----cgaatccagctcacttgc-t-aataagaatcattgacata--aatgtagt 559  
QY 2041 cgaatcattagcgtctgctcagagaggtgtgttcttctcactcagtaacagtaagtagct 2100  
DB 560 agatgctattagcgtctgctcagaga--gtgtgttcttctcactcagtagt-acaaagtagct 615  
QY 2101 agacaatgcttagaggtgtgttcttctcactcgtgttaggagcaaacacccat 2160  
DB 616 agacaatgctta-ggtgtgttcttctcactcgtgttaggagcaaacacccat--a 671  
QY 2161 ttccaatcctagaggaagcctcccccagcttgccttgccttcggcacaacatgctt 2220  
DB 672 ttccaatcctagaggaag-cctcccagcttgc-----tgtccttggcacaacatgct 726  
QY 2221 ctgagtaagtctgcaactcctccctggagagacataccacatcactctgtagtgc 2280  
DB 727 ctgagtaagt-tgacctaattcccttcggagacata--cgcatacactctgtagtgc 782  
QY 2281 gaggggagctaggaaggaagatccacacatcttcaataggttcaacagctacactcgtg 2340  
DB 783 gagggg--atgaagaagatataccacacatcttcaataggttcaacagctacactc--tg 837  
QY 2341 acaagtcagaatagggagcactgtctctatcccccacagtaggaagatcttgcgcaac 2400  
DB 838 acaagtcagaataggg--acaagtcctatcccccacagtagga--gagatctcggaaa--- 891  
QY 2401 ccccttcttctgaacaacagcccccagagcttgcacacatagctcctaaccacaaga 2460  
DB 892 -----ccttgaacaagccagagctt-gaaccttagctc-ecccaagaaga 936  
QY 2461 ctgagaaggaacatactcttcaagcttctcagagagcgtgtccttggaaatccaagaac 2520  
DB 937 ctgaga--gagacacatct-tctcagcttcttccagagagcgtgtc--tgggaatccaggaac 993  
QY 2521 gttttgactgcttaataaagccctgagactataaagtccatctatcgtggttttaac 2580  
DB 994 -ttttgactgcttaataaagccctgagactataaagtccatctatcgtggttttaac 1048  
QY 2581 tacaagtttttgaacatgtaggaagcagaacggtgccaagatgaataaaacatgacct 2640  
DB 1049 tacaagtttttga--atgttaggaagcagaaggtgccaagatgaataaaacatgacct 1104  
QY 2641 gtagaaggaaggaaggaaggaagaaactgtgtgtgtgtgtgtatcaataggaaggaagcact 2700  
DB 1105 gtagaaggaaggaaggaaggaagaaactgtgtgtgtgtgtgtatcaataggaaggaagcact 1164  
QY 2701 gggatcacctcgtgtcctttaggtccctctccatcagcagaaggaagcacttccaaag 2760  
DB 1165 gggatcacctc-ttcccttaggtccctctccatc-cagcagaaggaagcacttccaa- 1221  
QY 2761 tcatgctcctccgaagactgt 2820  
DB 1222 tcatgctcctccgaagactgt 1279  
QY 2821 gagccttaggttcaagttttgaataatgtgagacaacactgtcttggcagaaggtgtcgaaga 2880  
DB 1280 gagccttagg--tcaagtttgaataatgtgagacaacactgt--ctgtcagaaggtgtc--ga 1331  
QY 2881 gggagcctgtgtgtcaggaagtcaccagcgtccagcctccgggtgtgaagctcctgaggt 2940



Query Match	20.4%	Score 654.4	DB 22	Length 1667
Best Local Similarity	81.0%	Pred. No. 1.6e-171		
Matches 1431: Conservative	2	Mismatches 213	Indels 120	Gaps 53

OY	1441	gggaactcgcnaaagctgtgcatttgggtcctacgttgacctcgactctaaagacaatttca	1500
Db	1	gggaactcgcnaaagctgtgcatttgggtcctacgttgacctcgactctaaagacaatttca	59
OY	1501	gctgcctcactatttcttgggtggacacatgcgcgtgatttgcattctcgaagctgaatt	1560
Db	60	gctgcctcactc--atttgggtgacacatgcgcgtgatttgcattctcgaagctgaatt	117
OY	1561	gcattgatttgcacatagcaagatcnaataacaaagcgaaagcatactgaagacagagac	1620
Db	118	gcattgatttgcac--agcaagatcnaataacaaagcgaaagcatactgaagacagagac	174
OY	1621	ttgatctgcaagaagccttcaaatctcaaatcctacgtctgcacaaagcttccacactctg	1680
Db	175	ttgatctgcaagaagccttcaaatctcaaatcctacgtctgcacaaagcttccacactctg	232
OY	1681	gagcctaacggaagctcttccctcagctcagctgaagcttccagcctccaaagacagccctaat	1740
Db	233	gagcctaacggaagctcttccctcagctcagctgaagcttccagcctccaaagacagccctaat	290
OY	1741	gcaaaaatcccgtagtttcaagaacagacagcctgccccgggcctgcactatgaatcca	1800
Db	291	gcaaaaatccccc--tattcaagmacagacagcctgccccgggcctgcactatgaatcca	344
OY	1801	tcagaatgttgaaacccgcctctgcccccaacatactgatacatatctattctcagcgt	1860
Db	345	tcagaatgttgaaacccgcctctgcccccaacatactgatacatatctattctcagcgt	398
OY	1861	gtttagacaagaacgcctctggaagatgagacacagctaaagaactctcgctccggagtt	1920
Db	399	gtgtttgaaagaactgtgttggaagatgagacacagctaaagaactctcgctccggagtt	456
OY	1921	tctcttcacatctagacatactctgccagctcctcaactctgcaatgagcaagcttctgcaatgct	1980
Db	457	tctcttcacatctagacatactctgccagctcctctgcaatgagcaagcttctgcaatgct	512
OY	1981	ttgcaaacagacatccacgcgcactctgtctaaataaagaactctctgcatctaaatgtagct	2040
Db	513	ttgcaaacagacatccacgcgcactctgtctaaataaagaactctctgcatctaaatgtagct	558
OY	2041	cgatgctatttagcgtctgtgctcagagaggttggtttctctcaatcagtaacaaagtagct	2100
Db	559	agatgctatttagcgtctgtgctcagagaggttggtttctctcaatcagtaacaaagtagct	614
OY	2101	agacaatgcttaggggtgtgtttcttcaattcttctccctgctgaaggcaaacagcccat	2160
Db	615	agacaatgcttaggggtgtgtttcttcaattcttctccctgctgaaggcaaacagcccat	672
OY	2161	ttccaactctagaagaaagcctccccaagcctgctctgtctccctgggccaacactgtt	2220
Db	673	ttccaactctagaagaaagcctccccaagcctgctctgtctccctgggccaacactgtt	725
OY	2221	cttgagttaagtctgacctaaacttccctctgggaagacataccgcatcaactctgtgagctcc	2280
Db	726	cttgagttaagtctgacctaaacttccctctgggaagacataccgcatcaactctgtgagctcc	781
OY	2281	gagggtgatatgaaaggaataccaccacacttcttcataaggtctcaacagctacactctgtg	2340
Db	782	gagggtgatatgaaaggaataccaccaccacttcttcataaggtctcaacagctacactctg	836
OY	2341	acaagctcagaatagggaacacctctgtctctacccctccaaatggaaagatctctggccaac	2400
Db	837	acaagctcagaatagggtg--acaagctctctacccctccaaatggaaatctctggccaac	890
OY	2401	cccccttcttttgaanaacgcgcgccccagagctcttgcaacctagactccaaccagaaga	2460
Db	891	cccccttcttttgaanaacgcgcgccccagagctcttgcaacctagactccaaccagaaga	935
OY	2461	cttgaagaaggagacatacttcttaagctttttcagaagggcgctcttgggaatccaggaac	2520
Db	936	cttgaagaaggagacatactcctcagcttcttccagaagggcgctcttgggaatccaggaac	992

OY	2521	gtttttatgcttaattagaagcgcttgactataatgaatgctactatgcttgaacgtttaa	2580
Db	993	-ttttttatgcttaattagaagcgcttgactataatgaatgctactatgcttgaacgtttaa	1047
OY	2581	tacagctttttgaacatgcttagagagacagacgagcgacagagatgaaaaaacatgacctg	2640
Db	1048	tacagctttttgaac-atgcttagagagacagagcgagagag-aaaaaaacatgacctg	1103
OY	2641	gttaagaagaaagagacagaaagaaacttggttgagagatcaattagaagagagacact	2700
Db	1104	gttaagaagaaagagagacagaaagaaacttggttgagagatcaattagaagagagacact	1163
OY	2701	ggagttcacacttgcttctcttaggtgccctccctccatcagacagcaagaagacattctctaa	2760
Db	1164	ggagttcacacttc-ttctcttaggtgccctccctccatcagacagagagacattctcttaa	1220
OY	2761	tcatgacctcccgagaacgctggtcgtggagaaggttttaaaaaacaaaaatccagaagttaa	2820
Db	1221	tcatgacctcccgagaacgctggtcgtggagaaggtttt-aaaaacaaaaatccagaagt-ta	1278
OY	2821	gagccttagaggttcaagttttgaaaaattggagacaaacttgctcttgcaagaaggtgccaga	2880
Db	1279	gagccttagag-aa-tcaagtttgaaatctggagacaaactgtt-ctggcaaaaaggtgc--ga	1330
OY	2881	gcggagactctgtctcagagagatccacagcgctccagcctcgggtgttaaggtctctgaagt	2940
Db	1331	gcggagagcttggtgtctcagaggtt-cagagcgtccagcctcgggtgtt-aggttctgaagt	1388
OY	2941	gtgcgcatgggggagcctaagccttctctgtgtgacccgagagcttaagctgtgtgcacacacaca	3000
Db	1389	gtgcgcatctggggagcctaagccttctctgtgtga-cagaggtctcagcgtgtgpcacacacaca	1447
OY	3001	caaccacacacacacacacacacacacacaaattgggggcaacccacacatccacgttaaccagc	3060
Db	1448	caaccacacacacacacacacacacacacaaattgggggcaacccacatccacgttaacagc---	1503
OY	3061	tttaacacaaatgttatagtgctcccttttatcttaataagcctgtccctcttaaaagt	3120
Db	1504	--tttcaacaaatgttatagtgctcccttttatcttaataagcctgtccctcttaaaagt	1560
OY	3121	tattttattgtttaattttttctgtctcgtacgtttaattgtgaattgttaattgtcaataa	3180
Db	1561	tattttattgttatattttttctgtctcgtacgtttaattgtgaattgttaattgtcaataa	1620
OY	3181	gtgccttctgttagatggaaaaaaaa	3206
Db	1621	gtgccttctgttagatggaaaaaaaa	1646
RESULT 6			
AAF93580	AAF93580 standard; cDNA; 536 BP.		
XX	AC	AAF93580;	
XX	DT	21-MAY-2001 (first entry)	
XX	DE	Lung carcinoma cDNA encoding SRT protein SEQ ID 401.	
XX	KW	Human; SRT; gene therapy; gene mapping; tissue typing; ss.	
XX	OS	Homo sapiens.	
XX	PN	W0200107611-A2.	
XX	PD	01-FEB-2001.	
XX	PF	21-JUL-2000; 2000MO-US20006.	
XX	RR	26-JUL-1999; 99US-0145701.	
XX	PA	(GETH ) GENENTECH INC.	



PI Baker KP, Goddard A, Wood WI;  
 XX  
 DR WPI: 2001-112729/12.  
 XX

PT New isolated nucleic acid molecule encoding a SRT polypeptide is useful  
 PT for production of recombinant SRT polypeptides, gene mapping,  
 PT diagnosing genetic disorders and for gene therapy -  
 XX  
 PS

Claim 2; Fig 401; 663pp; English.

CC Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding  
 CC human SRT proteins. The cDNA sequences are isolated from various  
 CC different human tissue cDNA libraries. The invention relates to a method  
 CC for detecting cDNA encoding an SRT protein, a vector containing cDNA  
 CC encoding SRT, a host cell transformed with the vector, an isolated SRT  
 CC polypeptide, and an antibody which binds to SRT. The polynucleotide  
 CC sequence can be used in gene therapy and is useful in the recombinant  
 CC production of SRT polypeptides, as a hybridisation probe to screen  
 CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to  
 CC map the gene encoding the SRT polypeptides and analysing genetic  
 CC disorders, tissue typing and disease tissue detection. The SRT  
 CC polynucleotide sequences can be used in polymerase chain reaction,  
 CC screening for new therapeutic molecules and generation of antisense RNA  
 CC and DNA.  
 CC  
 XX

Sequence 536 BP; 166 A; 156 C; 77 G; 134 T; 3 other;

Query Match 11.4%; Score 365; DB 22; Length 536;  
 Best Local Similarity 90.5%; Pred. No. 2.6e-91;  
 Matches 503; Conservative 0; Mismatches 33; Indels 20; Gaps 10;

QY 279 taacacacacagaagaactgactagtcctctacagagtcgacgactatccactgaa 338  
 |||||||  
 Db 1 taacacacacagaagaactg-gantagtgctctacagtagcgacgactgataccactgaa 59  
 |||||||  
 QY 339 ctaattgcccctgaactgctagcaccacagaacataacactctctctccacagactact 398  
 |||||||  
 Db 60 ctaattt-ccttgaactgctagcaccacagaacataacactctctctccacagactact 118  
 |||||||  
 QY 399 tcactgtctcccccataataatagtaacataagttctctccacaaatctcactgctccc 458  
 |||||||  
 Db 119 tcactgtctcccccataataatag-aacatagttctctccacaaatctcactgctccc 177  
 |||||||  
 QY 459 cccataatagtaacataagttctctccacaaatctcactgctgacgacagtagag 518  
 |||||||  
 Db 178 cccataatagtaacataagttctctccacaaatctcactgctgacgacagtagag 237  
 |||||||  
 QY 519 tcaaccacaaatgtaaatcagtagtactcctctgacataatcacgcgttcactccaaa 578  
 |||||||  
 Db 238 tcaaccacaaatgtaaatc-ttagtacctctgacataatcacgcgttcactccaaa 296  
 |||||||  
 QY 579 tgatggttaattcaatagttctctctgaaacaaagtaacataatgaatgtcccca 638  
 |||||||  
 Db 297 tgatggtatga-tcaaatggtctctctgaaacaaagtaacataatgaatgtcccca 355  
 |||||||  
 QY 639 ccacagaagaacataacatccctcagtggtcctccactggacacggtttattgtagagacc 698  
 |||||||  
 Db 356 ccacagaagaacataacatcagc-ggctccacactgacacggtttattgtagagaccgc 414  
 |||||||  
 QY 699 atgcaagcctaaacagcagtgctccagcaatccttgccaagaatgtaacccctgtgca 758  
 |||||||  
 Db 415 a-----ccctaacaagcagag-gtccacagcaatccttgcca---agatgagccctgtgca 465  
 |||||||  
 QY 759 gataattggtatggttgtaagcttgctataataacagttttgctgtgttaagaag 818  
 |||||||  
 Db 466 gataattggtatggttgtaagctgc-----ataatacaagttttgctgtgttaagaag 520  
 |||||||  
 QY 819 ggtattactaacactc 834  
 |||||||  
 Db 521 ggtattactaacactc 536

RESULT 7  
 AAV87786  
 ID AAV87786 standard; cDNA: 313 BP.  
 XX  
 AC AAV87786;  
 XX  
 DT 12-FEB-1999 (first entry)  
 XX  
 DE EST clone ES206.  
 XX

KW Expressed sequence tag; secreted protein; haematopoiesis regulator;  
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;  
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolytics;  
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.  
 XX

OS Homo sapiens.  
 XX  
 PN W09845437-A2.  
 XX  
 PD 15-OCT-1998.  
 XX

PF 10-APR-1998; 98WO-US06956.  
 XX  
 PR 10-APR-1997; 97US-0837312.  
 XX

PA (GENEY ) GENETICS INST INC.  
 XX  
 PI Agostino MJ, Jacobs K, Lavalie ER, McCoy JM, Merberg D;  
 PI Racine LA, Spaulding V, Treacy M;  
 XX

DR WPI: 1999-070078/06.  
 XX  
 PT New polynucleotides encoding human secreted proteins - derived from  
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,  
 PT ovary, pituitary, retina and colon cDNA libraries  
 PT  
 XX

PS Claim 1; Page 179; 641pp; English.  
 XX  
 CC The present sequence represents an expressed sequence tag (EST), and is  
 CC a polynucleotide of the invention. The polynucleotides of the invention  
 CC are all secreted EST sequences isolated from a variety of human tissue  
 CC sources. The EST sequences and proteins encoded by them are predicted to  
 CC have useful biological activities which would make them suitable for  
 CC treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data is given. Suggested activities  
 CC include nutritional activity, immune stimulating or suppressing activity,  
 CC haematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition  
 CC activity. The EST sequences are also stated to be useful for gene  
 CC therapy.  
 CC  
 XX

Sequence 313 BP; 95 A; 103 C; 43 G; 72 T; 0 other;

Query Match 7.7%; Score 247.8; DB 20; Length 313;  
 Best Local Similarity 98.5%; Pred. No. 7.4e-59;  
 Matches 271; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 192 taaatgaagaagcattatcattactctcttgctgctctctctctgtaaacacagcca 251  
 |||||||  
 Db 37 taaatgaagaagcattatcattactctctctgctgctctctctctgtaaacacagcca 95  
 |||||||  
 QY 252 ccaacaaagcgaactcagctgctgtaacacacacagaagaactgagctagtgtccta 311  
 |||||||  
 Db 96 ccaacaaagcgaactcagctgctgtaacacacacagaagaactgagctagtgtccta 155  
 |||||||  
 QY 312 cagtagctgcagctgataccactgaactaatgtgcccctgaactgctgagcacaagca 371  
 |||||||  
 Db 156 cagtagctgcagctgataccactgaactaatlt-ccttgaactgctgagcacaagca 214  
 |||||||  
 QY 372 aatacacttcttccacacagctacttcaactgtctcccccataatgatacacatagt 431



```
XX PS Claim 2; Page 168; 472pp; English.
XX CC
XX CC The present invention describes colon tumour associated proteins (I) and
XX CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
XX CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
XX CC (II) may be used in the prevention, diagnosis and treatment of diseases
XX CC associated with inappropriate colon tumour associated protein (TCAP)
XX CC expression, such as colonic cancer. For example, (I) and (II) may be
XX CC used to treat disorders associated with decreased expression by
XX CC rectifying mutations or deletions in a patient's genome that affect the
XX CC activity of TCAPs by expressing inactive proteins or to supplement the
XX CC patients own production of them. Additionally, (II) may be used to
XX CC produce the TCAP proteins, by inserting the nucleic acids into a host
XX CC cell culturing the cell to express the protein. (II) and its
XX CC complementary sequences may also be used as DNA probes in diagnostic
XX CC polymerase chain reaction (PCR) and hybridisation assays to detect and
XX CC quantitate the presence of similar nucleic acids in samples, and
XX CC therefore which patients may be in need of restorative therapy. (I) may
XX CC also be used as antigens in the production of antibodies against TCAPs
XX CC and in assays to identify modulators of TCAP expression and activity.
XX CC Anti-(I) antibodies and antagonists may also be used to down regulate
XX CC TCAP expression and activity. The anti-(I) antibodies may also be used
XX CC as diagnostic agents for detecting the presence of TCAPs in samples
XX CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512
XX CC and AA24494 to AA24523 represent nucleotide and amino acid sequences
XX CC given in the exemplification of the present invention.
XX SQ
XX Sequence 417 BP; 87 A; 126 C; 88 G; 116 T; 0 other;
SQ
Query Match 6.2%; Score 200.2; DB 22; Length 417;
Best Local Similarity 88.3%; Pred. No. 1.5e-45;
Matches 356; Conservative 0; Mismatches 28; Indels 19; Gaps 12;
QY 2570 gggtttaattacagtttttgaaacatgtagagagcagagccagagagatgaa 2629
DB 394 GGGGTGACACTCTACAGTTTGGAA-ATGCTAGAGAGCAGAAAGGGCAGAGAG--TAA 339
QY 2630 aacatgacctgtagaagaagaagaagaagaagaagaagaagaagaagaaga 2689
DB 338 AACATGACCTGTGTAAGAAAGAGAGAGCAAAAGAAATAGTGGGAGATCAATTAGAG 279
QY 2690 agaggagacacccggagatcaccttcgttccttaggtccctccatcagcaagaagagc 2749
DB 278 AGGAGGACCTCGGAGATCACCTTC-TTCCTTAGAGTCCCTCTCTCAT-CAGCAAAAGAGC 221
QY 2750 acttctcttaagtcatgcctccgaagaagctgtaggaagaagtttaaaaaa 2809
DB 220 ACTTCTCTTAA-TCAATGCCCTCCCGAAGACTGCTGGAGAAAGTTT-AA 163
QY 2810 ccaagaagaagaagccttaggtcagtttgaataatggaacaaactgcttgagaa 2869
DB 162 CCAGAGAGT-AAAGAGCCTTAGG---TCACTTTGAATTTGCAACAAACT--GTCTGGCAAA 109
QY 2870 ggggtcccaagagcgagctgttcgtccagagagtcacagccgtccagcctcggtgtgaa 2929
DB 108 GGGTGC---GAGAGGGAGACTGTGCTCAGAGT--CCAGCCGTCCAGCCTCGGGGTGT-AG 54
QY 2930 gctctgaggtgtagcagctgagggcctcagccttctgtgtac 2972
DB 53 GTTCTGAGGTGTGTCATTTGGGCTCAGCCTTCCTCTGTGTGAC 11
RESULT 10
ID AAA77762 standard; cDNA; 401 BP.
XX AC AAA77762;
XX AC
XX DT 14-NOV-2000 (first entry)
XX DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:41.
```

```
XX KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
XX KW Immunotherapy; diagnosis; progression; ss.
XX OS Homo sapiens.
XX PN WO200037643-A2.
XX PD 29-JUN-2000.
XX PF 23-DEC-1999; 99WO-US30909.
XX PR 23-DEC-1998; 98US-0221298.
XX PR 02-JUL-1999; 99US-0347496.
XX PR 22-SEP-1999; 99US-0401064.
XX PR 19-NOV-1999; 99US-0444242.
XX PR 02-DEC-1999; 99US-0454150.
XX PA (CORI-) CORIXA CORP.
XX PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
XX PI Wang T, Yugu J;
XX DR WPI: 2000-442671/38.
XX PT New colon tumor polypeptides used to inhibit the development of cancer,
XX PT especially colon cancer, and for diagnosing and monitoring the
XX PT progression of the cancer -
XX PS
XX PS Claim 1; Page 99; 229pp; English.
XX CC Sequences AAA77722-A78199 represent 478 cDNAs encoding proteins or
XX CC portions of proteins which are associated with human colon tumours.
XX CC The invention also specifically discloses 8 human colon tumour proteins
XX CC (AAB1897-B11904). The nucleic acids, the polypeptides they encode, and
XX CC antigen presenting cells (APCs, preferably dendritic cells) expressing
XX CC such polypeptides may be used in vaccines that target tumour cells,
XX CC especially colon tumour cells, thereby inhibiting the development of
XX CC cancer. T-cells specific for the polypeptide expressed by the APC are
XX CC used to remove tumour cells from biological samples, especially blood or
XX CC fractions thereof. The sample or the isolated T-cells specific for the
XX CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
XX CC CD8+ T-cells from a patient may be incubated with a polypeptide or
XX CC nucleic acid of the invention, or an APC expressing such a polypeptide,
XX CC to cause the proliferation of specific T-cells. The T-cells can be
XX CC cloned and then administered back to the patient to inhibit cancer
XX CC development. Nucleic acids encoding the polypeptides and antibodies
XX CC against the polypeptides may be used to determine the expression level
XX CC of a tumour protein of the invention, and therefore to determine whether
XX CC cancer cells are present. Such diagnostic methods may also be used to
XX CC monitor the progression of a cancer by repeating the processes at time
XX CC intervals, and comparing the current result to previous results. The
XX CC present sequence represents a cDNA encoding a human colon tumour
XX CC polypeptide.
XX SQ
XX Sequence 401 BP; 113 A; 85 C; 114 G; 85 T; 4 other;
SQ
Query Match 5.9%; Score 187.6; DB 21; Length 401;
Best Local Similarity 85.7%; Pred. No. 4.6e-42;
Matches 336; Conservative 0; Mismatches 38; Indels 18; Gaps 11;
QY 2570 gggtttaattacagtttttgaaacatgctgagggcagagggccagagagtataaa 2629
DB 23 ggggtgcactctcaagtttttgaa-atgctagagagcagaaggcagagag---taaa 78
QY 2630 aacatgacctgtagaagaagaagaagaagaagaagaagaagaagaagaagaaga 2689
DB 79 aacatgacctgtagaagaagaagaagaagaagaagaagaagaagaagaagaaga 138
QY 2690 agaggagacactgtagatcaccttcgttccttaggtccctccatcagcagaaagagc 2749
DB 139 agaggagacactgtagatcaccttc-ttccttaggtccctccatcagcagaaagagc 196
```





DT 14-JUN-2000 (first entry)  
XX Human colon cancer differentially expressed nucleotide sequence #132.  
DE  
XX  
KM Colon cancer: detect; differential expression; human; treatment;  
KM detect mutation; non-invasive diagnostic method; ds.  
OS Homo sapiens.  
XX  
XX WO200012702-A2.  
XX  
XX 09-MAR-2000.  
XX  
XX 30-AUG-1999; 99WO-US19424.  
XX  
XX 31-AUG-1998; 98US-0098639.  
XX 27-JAN-1999; 99US-0117393.  
XX  
XX (FARB ) BAYER CORP.  
XX  
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;  
PI Catino TU, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;  
PI Schlegel R;  
XX  
XX WPI; 2000-256641/22.  
XX  
XX Novel nucleic acids and proteins for identifying therapeutic agents  
XX useful for treating and diagnosing cancer, especially colon cancer  
XX  
PS Claim 16; Page 177; 345pp; English.  
XX  
XX This sequence represents a human nucleotide sequence which is  
XX differentially expressed in colon cancer cells compared to the expression  
XX levels in normal cells. The nucleotide sequence can be used as a source  
XX of primers and probes. The nucleotide sequence is useful for determining  
XX the phenotype of a cell by detecting the differential expression of the  
XX sequence relative to a normal cell. The probes derived from the sequence  
XX can also be used to determine the phenotype of cells in a sample. Probes  
XX and antibodies which hybridise to the nucleotide sequence can also be  
XX used to determine the phenotype of a cell. The primers are useful for  
XX detecting a mutation in a test nucleotide sequence and also for detecting  
XX cancer, preferably colon cancer. Antibodies against the protein encoded  
XX by the nucleotide sequence can also be used in a method to detect colon  
XX cancer. The diagnostic method is non-invasive and accurate for diagnosing  
XX colon cancer at an early stage.  
XX  
SO Sequence 593 BP; 173 A; 131 C; 137 G; 142 T; 10 other;  
Query Match 3.88; Score 123.2; DB 21; Length 593;  
Best Local Similarity 75.6%; Pred. No. 4.6e-24;  
Matches 437; Conservative 0; Mismatches 87; Indels 54; Gaps 21;  
QY 2097 actgagacatgcttaggggtgttcttaattcttccctgtgtggaacaagacc 2156  
DB 1 acgagacaaaggcta-gggctgttcttcttaattcttccctgtgtggaacaagacc 58  
QY 2157 ccatctcaaatctagagaaagcctcccgacatgcttgcgccctggcgcaaacat 2216  
DB 59 cattcc--aatctagagaaag--ctcccgacatgcttgcctctggc-----aaacat 111  
QY 2217 gctctctgagttaagtacctaactccctggagacgacatacagcatcaactgtgag 2276  
DB 112 tgcctctgagttaag-tgacctaatccctggagacata--cgatcaactgtgag 167  
QY 2277 gtccgaggggagatgagaaaggaataccacatcttccataggtgacaaagctacatc 2336  
DB 168 gtccgagggg--atgagaaaggatgacccacacatcttcaaggtcacaagctcacttc 224  
QY 2337 cgtgacagatcagaatagggacacactgctctatccctccaaaggagagatctggcc 2396  
DB 225 --tgacaagtcagaatagg--acactgcttctatccctccaaatgga-gagatctggca 279

QY 2397 aaacccccccttlttgaanaaccagcccccagagcttggcaactagctcaaccag 2456  
DB 280 accctt-----gaacagccagagctt-gaacactagcttc-accacag 321  
QY 2457 aagactggaagagacatatctttagcttcttcaggagggctgtggtggaatcag 2516  
DB 322 aagactggaag-gagacataac-tctagcttcttcaggagggctgtgc-tgggaatcag 378  
QY 2517 gaactgttgaatcgaataggaagccttgactataataatgctcctatgggttc 2576  
DB 379 gaac-tttttagatcattataggaagccttgactataataatgctcctatgggttc 432  
QY 2577 aatctacagttttgaaacatgctagagagcagaacgggccaagagataaacaatga 2636  
DB 433 actctacagttttgaa-atgctagagagcagaagggc-----aaaataaacaatga 486  
QY 2637 cctgtgtagaaggaaggaagcagaacactggtggg 2674  
DB 487 cctgtgtgaaaggaanaagaaagaacttgggggg 524  
RESULT 15  
AAA16148  
ID AAA16148 standard; DNA: 620 BP.  
XX  
XX AAA16148;  
XX  
XX 14-JUN-2000 (first entry)  
XX  
XX  
DE Human colon cancer differentially expressed nucleotide sequence #153.  
XX  
XX Colon cancer: detect; differential expression; human; treatment;  
XX detect mutation; non-invasive diagnostic method; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200012702-A2.  
XX  
XX 09-MAR-2000.  
XX  
XX 30-AUG-1999; 99WO-US19424.  
XX  
XX 31-AUG-1998; 98US-0098639.  
XX 27-JAN-1999; 99US-0117393.  
XX  
XX (FARB ) BAYER CORP.  
XX  
PI Endege WO, Steinmann KE, Astle JH, Burgess CC, Carroll E;  
PI Catino TU, Dwivedi P, Ford DM, Lewis ME, Molino GA, Monahan JE;  
PI Schlegel R;  
XX  
XX WPI; 2000-256641/22.  
XX  
XX Novel nucleic acids and proteins for identifying therapeutic agents  
XX useful for treating and diagnosing cancer, especially colon cancer  
XX  
PS Claim 16; Page 185; 345pp; English.  
XX  
XX This sequence represents a human nucleotide sequence which is  
XX differentially expressed in colon cancer cells compared to the expression  
XX levels in normal cells. The nucleotide sequence can be used as a source  
XX of primers and probes. The nucleotide sequence is useful for determining  
XX the phenotype of a cell by detecting the differential expression of the  
XX sequence relative to a normal cell. The probes derived from the sequence  
XX can also be used to determine the phenotype of cells in a sample. Probes  
XX and antibodies which hybridise to the nucleotide sequence can also be  
XX used to determine the phenotype of a cell. The primers are useful for  
XX detecting a mutation in a test nucleotide sequence and also for detecting  
XX cancer, preferably colon cancer. Antibodies against the protein encoded  
XX by the nucleotide sequence can also be used in a method to detect colon  
XX cancer. The diagnostic method is non-invasive and accurate for diagnosing  
XX colon cancer at an early stage.  
XX



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:24:16 ; Search time 8456.85 seconds  
(without alignments)  
6254.097 Million cell updates/sec

Title: US-09-867-034-4  
Perfect score: 3206  
Sequence: 1 ttccgctcgagtgtaaac.....ttgtatgatcgaataaaaaa 3206

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_ba:\*  
2: gb\_hcg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_om:\*  
20: em\_ov:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hugo\_hum:\*  
31: em\_hugo\_inv:\*  
32: em\_htg\_rod:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1466	45.7	2887	9 AF286113	AF286113 Homo sapi
2	1458.8	45.5	2864	9 AB035807	AB035807 Homo sapi
3	1030	32.1	2265	6 AX193124	AX193124 Sequence
4	1030	32.1	2265	9 AK000070	AK000070 Homo sapi
5	1009	31.5	2263	6 AX011684	AX011684 Sequence
6	462.4	14.4	186893	2 AC026342	AC026342 Homo sapi
7	365	11.4	536	6 AX079657	AX079657 Sequence
8	200.2	6.2	417	6 AX192560	AX192560 Sequence
9	187.6	5.9	401	6 AX192474	AX192474 Sequence
10	182	5.7	118407	2 AC026629	AC026629 Homo sapi
11	160.2	5.0	2636	10 RN089744	RN089744
12	149.6	4.7	1325	9 HSM801289	HSM801289
13	149.6	4.7	5182	9 AB023198	AB023198 Homo sapi
14	149.6	4.7	232816	2 AC012362	AC012362 Homo sapi
15	144	4.5	2968	10 MUSANT10A	MUSANT10A
16	136.8	4.3	294	11 G21920	G21920 human STS W
17	102.6	3.2	186893	2 AC026342	AC026342 Homo sapi
18	61.4	1.9	99996	2 AL592077	AL592077 Dario rer
19	60.6	1.9	14091	3 SUSMP1	SUSMP1
20	60.6	1.9	22398	5 FRU271723	FRU271723 Fugu rubr
21	59.8	1.9	155550	2 AC020227	AC020227 Drosophi
22	59.8	1.9	161278	3 AC009749	AC009749 Drosophi
23	59.8	1.9	241429	3 AE003608	AE003608 Drosophi
24	58	1.8	1470	3 LSI132603	LSI132603 Litomosol
25	57.8	1.8	2161	5 XELFTMCIX	XELFTMCIX
26	56.2	1.8	79433	2 AL591180	AL591180 Dario rer
27	55.2	1.7	186222	2 AC020858	AC020858 Mus muscu
28	55	1.7	8438	14 AF160185	AF160185 Fowl aden
29	54.6	1.7	379	8 AF286605	AF286605 Pirus tae
30	54.6	1.7	185994	2 AC002042	AC002042 Homo sapi
31	54.6	1.7	327209	3 AE003575	AE003575 Drosophi
32	54.4	1.7	172307	2 AC044842	AC044842 Homo sapi
33	54	1.7	227724	2 AF336381	AF336381 Mus muscu
34	53.8	1.7	64998	2 AC014035	AC014035 Drosophi
35	53.8	1.7	269518	3 AE003751	AE003751 Drosophi
36	53	1.7	5163	6 AR097042	AR097042 Sequence
37	53	1.7	5318	6 AR097041	AR097041 Sequence
38	53	1.7	182978	2 AC021176	AC021176 Homo sapi
39	52.6	1.6	158414	10 AC005960	AC005960 Mus muscu
40	52	1.6	448	11 G45188	G45188
41	51.8	1.6	177989	2 AL591430	AL591430 Mus muscu
42	51.8	1.6	200125	2 AL591064	AL591064 Mus muscu
43	51.4	1.6	7218	6 I66494	I66494 Sequence 14
44	51.4	1.6	25666	2 AC020076	AC020076 Drosophi
45	51.4	1.6	323461	3 AE003491	AE003491 Drosophi

#### ALIGNMENTS

RESULT	1	LOCUS	AF286113	DEFINITION	Homo sapiens 2887 bp mRNA	PR1	27-MAY-2001
ACCESSION	AF286113	VERSION	AF286113.1	GI:14209831			
KEYWORDS		SOURCE	human.				
ORGANISM			Homo sapiens				
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS			Williams,S.J., Wreschner,D.H., Tran,M., Eyre,H.J., Sutherland,G.R. and McGuckin,M.A.				
TITLE			MUC13, a Novel Human Cell Surface Mucin Expressed by Epithelial and Hemopoietic Cells				
JOURNAL			J. Biol. Chem. 276 (21), 18327-18336 (2001)				
REFERENCE			2 (bases 1 to 2887)				
AUTHORS			Williams,S.J., Wreschner,D.H., Tran,M., Eyre,H.J., Sutherland,G.R.				



QY	1789	tattagaatccatcaaatgttggaacccgcgaatggcccccaacatagtataatcat	1848
Db	1561	TATTAGAAAT-CATPAGAAATGTGGAAACCGCATAGGCCCCCAACCAATGTACAGCTATT	1619
QY	1849	tattctgacagttcttttgacaagaacatgcatgtgaagtagacacaagctaaagactctgg	1908
Db	1620	TTT-----AGAGTGTTTAAAGAACTGATGTGGAAGTAGACACCA--GTAAAGATCTGG	1671
QY	1909	ccctcggaagttctctcttcctcatctagacatactgccagtcctcatctgcaatggaacg	1968
Db	1672	CCCTCGGGGTTTTTCTTCATCTGACATCTGCCAGCCTCTCGAATGTGAATGTGCAATG	1731
QY	1969	ttgtagcaatgtcttgcaagaacatcccaagctcaactgtgtaaatagaatctatgaat	2028
Db	1732	TTTTGCA-----CGAATTCAGCTCACTTGGCT-AAATPAGAAATCTATGACAT	1776
QY	2029	taaatgttagctcgatgcatctatagcgcgtgtgctcaagaagtgaggttctctcaactgt	2088
Db	1777	TA--AAATGTATGTATGCTTATGTAGCCCTGTGTAGAGA---GGTGGTTTTCTTAATCACT	1831
QY	2089	aacaagaactgtagacaatgtcttaagggtgtgttcttaatctcttccctgttagagca	2148
Db	1832	-ACAAATAGTGTAGACAAATGTGA--GGGTGTGTTTTCTTAACTCTTT--CCTGTATGGCA	1888
QY	2149	acaaagaccatcttccaatctatgagaagaagctccccaagatctgtgtctcctgagc	2208
Db	1889	ACAAAGAACCAATTTCC--AACTATGAGAGAAAG-CTCCCCAGCAATTCCTGTCTCTGGGC--	1943
QY	2209	caaaccaatgctctcttgaatgaatgtaaccttaactcccttggaagacataccagatca	2268
Db	1944	--AAACTTTCCTCTTGAGTTAAG-TCACTTAATTCCTCTGGAGACATTA--CGATATA	1997
QY	2269	ctgtgtagtgcgagagggagatgagaagaagatataccacaatcttcatatgagtlcaaac	2328
Db	1998	CTGTGAGGTCGCGAGGGG--ATGAGAAAGGATATCCACCACTTTCAAGGGTCACAACCT	2055
QY	2329	tacactctgtgacaagtcagaatagaagggaacactgtctctatccctccaatgtagaaga	2388
Db	2056	CACCTCT--TGACAACTGTGAATTAAGG--ACACTGCTCTTATTCCTCTCAATGGA-GAGA	2110
QY	2389	ttcgtgccaaccccccttttttgaaaacagcccccaagagcttggaaactgagctc	2448
Db	2110	TTCTGGCA-----CCTTTGAACAGCCGAGAGCTT-GCMACTATGCTCT	2152
QY	2449	aaacccaagaagacttggaaaagagacatacttcaagcttcttcaagagcgctgcttgg	2508
Db	2153	-ACCAAGAAAGACTGGAAA-GAACAATATC-TCACAGCTTTTTCAGAGAGCGCTGCC-TGG	2208
QY	2509	gaatccaagaaagttttgtatgcttaattgaagagcctgtgactataatagtcatctat	2568
Db	2209	GAAATCCAGGAAC-TTTTTGATGCTAATTTAGAAAGCGCTGACATAAAATGCTCCAC---TA	2263
QY	2569	ggggttttaacttaacagtttttgaacatcttgaagagcggaacggggccagagagttaa	2628
Db	2264	TGGGTCACCTCTTACATTTTTTAA-ATGCTAGAGAGCAAGAAAGGGGCAAGAG--TAA	2319
QY	2629	aaacatgacctgttagaaggaagaagagcaagaagaactgggttgaggagatcaataga	2688
Db	2320	AAACATGACCTGTGTAGAGAAAGAAAGAGGCAAGAAAGAACTGGGTGGGAGATCAATTGA	2379
QY	2689	gaagaagcaccgtggatccaactgtctccttagtctccctctccatgtcagcaagaag	2748
Db	2380	GAGGAGGACACTGTGGATCCACTTC-TTCCCTTAGGTCCTCTCCAT-CAGCAAAAGAG	2437
QY	2749	caactctctaagtcataagccctcccggaagctgtgtgaggaaagtttaaaaaaanaa	2808
Db	2438	CACCTCTCTTAA-TCATATCCCTCCGGAAGACTGTGGGGAAGGTTT-AAAAACAAAAA	2495
QY	2809	tccagagataaagaccttggagtgcaagtgttgaaatgtgaacaacactgtctgtgcaa	2868
Db	2496	TCCGAGAGAT-AAAGAGCTTAGGTGACG---TTTGAAATTTGAGAACAAACTGT--CTGGCA	2549
QY	2869	agggtgcgaagcggaagcttgttctcaggaagttcccaagcgttcaagctctggggttaa	2928

Db	Accession	Source	Organism	Reference	Authors	Title	Journal	Reference	Authors	Title	Journal
Db	2550	AGG6TC---	GAGACGGAGCTTGTCTCAGAGACT	CCAGCCGCTCCAGCTCGGGGTG	A	2604					
Qy	2929	ggtctctgaggtgltgcacatg	gggggagcctcagcctctctg	ltgaccgaggtcagctgtg		2988					
Db	2605	GGTTCTGTAGGAGTGTCCAT	TGGGGCTCAGCCTTCTGTGTA	-CAGAGGCTCACACTGTG		2663					
Qy	2989	ggcaccacaacacacacacac	acacacacacacacacacac	acacacacacacacacacac		3048					
Db	2664	GCCACCAACACACACACAC	ACACACACACACACACAC	ACACACACACACACACAC		2723					
Qy	3049	acgtaccacagcttaacacaa	tttttttttttttttttttt	tttttttttttttttttttt		3108					
Db	2724	GTACAGC-----	TTTTACAAATCTTATATATGTC	TTTTTATTTCTTAAT-GCCTTGT		2776					
Qy	3109	ccctctaaagtattattatt	tttttttttttttttttttt	tttttttttttttttttttt		3168					
Db	2777	CCTCTAAAGTATTATTATT	TGTTATTATTATTATTATT	TGTTGTTCTTACACTGTTAAT		2836					
Qy	3169	taatgcaataaagtcctt	gtttagatggaacacacac	acacacacacacacacac		3206					
Db	2837	TAATGCAATAAAGTGCCT	TTGTGTATCATGTGTA	GAACAAA 2874							
RESULT	2										
LOCUS	AB035807	2864 bp	mRNA	PRI	19-DEC-2000						
DEFINITION		Homo sapiens mRNA for RECC, complete cds.									
ACCESSION	AB035807										
VERSION	AB035807.1	GI:11907512									
KEYWORDS		RECC.									
SOURCE		Homo sapiens cDNA to mRNA.									
ORGANISM		Homo sapiens									
REFERENCE		Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
AUTHORS		1 (sites)									
TITLE		Masaoshi, I., Furukawa, Y., Akashi, H., Han, H., Nakajima, Y., Sugano, S., Ogawa, M. and Nakamura, Y.									
JOURNAL		Isolation and characterization of a novel human gene RECC encoding a mucin-like glycoprotein, homologous to murine cell surface antigen 114/Al0, and its reduced expression in colorectal cancers									
REFERENCE		Unpublished									
AUTHORS		2 (bases 1 to 2864)									
TITLE		Nakamura, Y.									
JOURNAL		Direct Submission									
REFERENCE		Submitted (11-DEC-1999) to the DDBJ/EMBL/GenBank databases. Yusuke Nakamura, Institute of Medical Science, The University of Tokyo, Laboratory of Molecular Medicine, Human Genome Center; 4-6-1, Shirokaneai, Minato-ku, Tokyo 108-8639, Japan (E-mail: furukawa@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5375, Fax: 81-3-5449-5433)									
FEATURES		Location/Qualifiers									
SOURCE		1. 2864									
gene		/organism="Homo sapiens"									
CDS		/db_xref="taxon:9606"									
		/chromosome="3"									
		/map="3q21"									
		25..1563									
		/gene="RECC"									
		25..1563									
		/gene="RECC"									
		/codon_start=1									
		/product="RECC"									
		/protein_id="BAB19651.1"									
		/db_xref="GI:11907513"									
		/translation="MKAIIHTLLALISVNTAINOGNSADAVTTEETATSGPTVAAD									
		TTEENPEETASTANTSPFPATSPAPPIISTHSSSTIPPPAPPIISTHSSSTIPPTG									

TKHIEENLIDEDFONLKRSTGFTNLGABGVFPVRVITTAHDSOMNPNYRHSMP  
BASE COUNT 854 a 670 c 611 g 729 t  
ORIGIN  
Query Match 45.5%; Score 1458.8; DB 9; Length 2864;  
Best Local Similarity 85.0%; Pred. No. 0;  
Matches 2555; Conservative 0; Mismatches 287; Indels 164; Gaps 76;  
QY 192 taaatgaaagcatatctatctatcttctgtgctgcttcccttctgtaaacacaca 251  
DB 21 TAAATGAAAGCCATCATATCTACTCTCTGTC-TCTCTCTCTGTAACACAGCCA 79  
QY 252 ccaaccaaggcaactcagctgtagctgtaaacacacacagaactgagactagtgctca 311  
DB 80 CCAACCAAGCACTCAGTGATGCTGTAAACACAGAAAGTGGACATAGTGCTCTA 139  
QY 312 caagtactgagctgatacacaactgaactaattgcccctgaactgctagcacacaaga 371  
DB 140 CAGTAGCTGACGTGATACACAGTAACCTAATTT-CCCTAAACTGCTACACACACAGCA 198  
QY 372 aataacactcttcccaacagacttaactcctgctcccccataatctagtaacatagt 431  
DB 199 AATACACTCTCTCCCAACAGCTACTTCACTGCTCCCCCATTAATTAATAGCATAGT 258  
QY 432 tctccacaactcctacacactgctcccccataaattagtaacatagtctccacacat 491  
DB 259 TCCCTCCCAATTCCTACACCTGCTCCCCCATTAATTAATTAATTAATTAATTAAT 318  
QY 492 cctatacctactgctgagacagtgagtaacacacaaatgaattgaattagctactc 551  
DB 319 CCAATACCTACTGCTGACAGAGTGAAGTCAACCAATGTAATTAATTAATTAATTAAT 377  
QY 552 tgaataataacgcctcactcactcacaatgtagtaatactacacatggttccctcga 611  
DB 378 TGACATTAATACCGCTTCACTCATCTCAAAATGATGATTAA-TCACATAGGTTCTCTTCTA 436  
QY 612 cacaagaataacatgaaatgctcccccacacagagaacaaatcactcagtgagctccc 671  
DB 437 CACAAATGAATGAATGATGCTCCACACAGAGACAAATCAATCAATCAATCAATCAAT 495  
QY 672 actgagacgcgttattttagtagacatgcaagccttaaacagacagtgctccagaat 731  
DB 496 ACTGACACCGCTTATTTGAGACACGCA----CCCTAAACAGACACAG-CTCCACAGCAAT 549  
QY 732 ccttgcaagaatgacccctgctgacataatcgttatctgttctgtaacgtgcata 791  
DB 550 CATTGCCA---AGATGATCCCTGTGACAGATTAATTCGTTATGTGTTAAGCTGC-----ATA 601  
QY 792 atacaagtttctgctgctgttttagaaggatatactacaactcttctacatgtaagaag 851  
DB 602 ATACAAATTTTCTGCTGTGTTGAAAGGATTAATCTCAACTCTTCTCACTCAATGTAAGAAG 661  
QY 852 gaagagatccctcctggaagattttagtagacatgatacagaacaaatttgccagaagag 911  
DB 662 GAAAGGATTTCCCTGG-GAAGATTTCAGTACATATCAGAAACATTTGACCCGAGAGAG 720  
QY 912 aaacattccaatggcctcaatcaagactgcatagtgaaatctactagctgtttaaagaagta 971  
DB 721 AAACATTCCATGGCCTTATCAAGACTTGCATAGTGAATTAATTAATTAATTAATTAATTA 780  
QY 972 ttgagacatctgtttagtagacagactgtatcttactctgtaagagacatctctgtaacc 1031  
DB 781 TTTGGCAGATCTGTTTATGAGACAGACTGTAAATCTTACTGTAA-GCACATCTCTGTCAAC 839  
QY 1032 aagatcttgaatagcgctgctgtagacaagtttctgttaatgtaacaatagtaacaatttc 1091  
DB 840 AAAA-TCGTGAATGCGGTC-TGATGACAAAG-TTGTGTTAAATGAACAATTAATTAATTTT 896  
QY 1092 gggcagaacacacagtgacaatgagaagactgtagctgaggaagaatgaatgaacaattt 1151  
DB 897 GGCAGAAACCAAGTACATATGAGAAAGACTGTGACT-GAGAAATTAATTAAGCAA-TT 954

QY 1152 ataagtagctcaagcaacttcttaactatgatactgagccctgctcggtgtagatg 1211  
DB 955 AGAAGTAGCTCAAGCAAC-TTCTAAACATATGATTGACCC-----TTGCGTGATTA 1007  
QY 1212 agggctgggaacacagactgctgtagatgactgctcaatgagtggttagtagagatgca 1271  
DB 1008 TTAATGCTGTAAACAGACTC--CGGATGACTGCTCAAT-GGTTTATGACATGCA--TTGC 1062  
QY 1272 aatgtagctgcaaaagccttaacccacagagcccttctgctgcttccacgtccaga 1331  
DB 1063 AAATCTGACTGCAAGAGCCTTAACCCACAGAGCCCTTTGCGTGTGCTTCAATCTCA-A 1121  
QY 1332 gctgctgtagctgcaacgcacagacacaaagcgaatgcttaataaagaagagtggtggg 1391  
DB 1122 GTGTCTGATGCTGCAACGACACAGCAAGC-AATGCTTAATTAAGAAAGAGTGTTGGG 1180  
QY 1392 tcccctgcaatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1451  
DB 1181 CCCCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1235  
QY 1452 aagatgcatcttgggctacagtgagactgactgactgaagacaatctcaactgactcca 1511  
DB 1236 AAGTGTGATTT-GGCTACAGTGAGACTGATGATGAAGCAATTAATTAATTAATTAATTA 1294  
QY 1512 cttaattgtagac 1571  
DB 1295 CT--ATTGTGGACACATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1352  
QY 1572 caactagaagaataaataaagaagaacacacacacacacacacacacacacacacacac 1631  
DB 1353 CAC-AGCAAGATCAAAATTAACAAAA-CGAAACATATTAAGAA-GAGAACTTGAATGACGA 1409  
QY 1632 agacttcaaaactcaaaactcagctgctgcaacacacacacacacacacacacacacac 1691  
DB 1410 AGACTTCAAAATCTAAACTGCGGCTG-ACAGCTTACACATCT-TGAGACAGAAAGG 1467  
QY 1692 agcgtctccctcagtgagtagtaacgctcacaagagacgcctagatgaacaaatccc 1751  
DB 1468 AGCGTCTTTCCTTAAGGTCAGAGTACGCGCTCCAGAGA--CAACCGAGATCAAAATCCCT 1525  
QY 1752 gtagattcaagaacagac 1811  
DB 1526 A----TTCAAGACACACACACATG-CCCGCCCTGACTTATTAAT-CATTAACAAATGAG 1579  
QY 1812 aaccgcgcatgcccccaac 1871  
DB 1580 AACCGGCATGGCCCCAACCAATGTAACAGCTAATTAAT-----AGAGTGTTAAGAA 1633  
QY 1872 actgcatgagaagtgagac 1931  
DB 1634 GACTGATGGAGAGTGAAGTGAACCA--GTAAGATCTGCGCTCCGCGGCTTCTTCCATCT 1691  
QY 1932 ctgacatactgcaagctcctcactgcaatgcaacgcttctgcaatgctctgcaacgac 1991  
DB 1692 GACATCTGCCAGCCTCTCTCAATGGAAGTTGTGAATGTTTGCA-----CG 1737  
QY 1992 atccagctcacttgcataaataagaatcatalgacatlaacatgtagctgactgactata 2051  
DB 1738 AATTCAGCTACTTTCGT-AAATTAAGATCTATGACATTA--AATGTAATGATGCTATTA 1794  
QY 2052 ggcctgctgctcagaagtggttcttcaatcagtaacaaagtaactgagacaatgctc 2111  
DB 1795 GCGCTTGTCAAGAGAGT--GGTTTCTTCAATCAAGT-ACAAAAGTACTGAGCAATAGGTT 1850  
QY 2112 aggggttgggttcttaattcttccctgtaggaagaagaagcccatctcaaatctca 2171  
DB 1851 A-GGGTGTGTTCTTAATCTTTT-CCGTGTAAGGCAACAAAGCAATTTCC--AATCTA 1906  
QY 2172 gaggaaagcctcccaagcatgcttgcctccctgggccaacacatgcttctgagtaag 2231  
DB 1907 GAGGAAG-CTCCCCAGCATTCCTGCTGCTGCGC-----AAACATTTGCTTGAAGTTAAG 1961

2232 ttgaacttaactcccttggaagacatcacatcaactgtggaaggtccgagggatga 2291  
1962 -TGACCTAATTCCTCCCTGGGACATA---CGCATCACTGTGGAGTCCGAGGGG--ATG 2015  
2292 gaaaggatccccaccatcttcataaggtlcaaaagctacactcgtgcaagtcagaa 2351  
2016 AGAAGGGATACCCACACACTTTCAGAGGCTCAACAGCTCACTCTC---TGACAAAGTCAGAA 2072  
2352 taagggaacactgtcttcatccctcccaatggagagatctggcccaacccccctttt 2411  
2073 TAGGG--ACACTGCTCTATCCCTCCAAATGGA--GAGATTCTGGCA----- 2115  
2412 tgaaaaccagggcccccagagcttgcaaccctagcctcaaccacaaagacgtgaaaggag 2471  
2116 --CCTTTGAACAGCCCAAGCTT--GCACCTAGCCTC--ACCCAGAAACATCTGGAAA--GAG 2170  
2472 acatacttctcaagcttcttcaagagagcgctgcttgggaatccaggaacgttttgatgc 2531  
2171 ACATATC--TCTACGCTTTTTCAGAGGCGGTGCC--TGGGAATCCAGGAAC--TTTTGATGC 2227  
2532 taattagaagggcctggagactaataatgtcatcatatgaggtttaaactcaagttttg 2591  
2228 TAATTAGAGGCTGTGAGCTAATAATGTCCAC---TATGGGTGCACTCTACAGTTTWTG 2283  
2592 aacatgtctagaagcagaacgagcagagagataaaacatgacctcgtgtagaagag 2651  
2284 AA-ATGCTAGAGAGCAGAGAGGGCGAGAGAG---TAAAAACATGACCTGTGTAAGAAAGAG 2339  
2652 agagacaaaggaacactgggtggaagatcaatagaagagagacactggatcacct 2711  
2340 AGAGGCAAAAGGAACTGGGTGGGAGATCAATTAGAGAGAGGACCTGGGATCCACT 2399  
2712 tctgttcccttagtcccccctccatgacagcaaaagagacacttcttaagttaagctccc 2771  
2400 TC-TTCTTAGTCCCTCCCTCCCTCAT--CAGCAAAAGAGCACTTCTTA--TATGCTCCCTC 2456  
2772 cgaagactgcttggaagagtttaaaaaaacaataaccaggaagtaagagccttagag 2831  
2457 CGAAGACTGGCTGGGAGAAAGTTT--AAAAACAAAATCCAGAGAT--AAGAGCCTTAGG-- 2513  
2832 tcaagtttgaaaatgtgagacaaactgtctgtgcaaaaggttgccaagagcgttctgt 2891  
2514 --TCAGTTTGAATTTGGAGACAAACT--GTCGTGGCAAAAGGGTGC---GAGAGGGAGCTTG 2566  
2892 tgcctcaagagtcceagccgctcagcctcgggtgtgaaggtcttcaaggttgcacatg 2951  
2567 TGCCTCAGAGATC--CAGCCGCTCAGCCTCGGGGTGT--AGGTTTCTGAGGTGTGCCATTGGG 2624  
2952 gctcagcctctctgtgtgacccgaggtcagctgtgtgccaaccaacaacaacacacac 3011  
2625 GCTCAGGCTTCTGTGTGGA--CAGAGGCTCAGCTGTGGCCACACCAACACACACACAC 2683  
3012 acacaaaccacacacaaatggtgggcaaacacacatcaccgtaaccaagcttlaacacaa 3071  
2684 ACACAAACACACACAAATGGGGCCACACATCCACTACAAAGC-----TTTTACAAA 2737  
3072 tgtattatgtcccttttatttcttaataagccctgtcctctttaaagttttattttt 3131  
2738 TGTATTATGTCTCTTTTATTATTCTAAT--GCTTGTCTCTTAATAAATATTATTATTG 2796  
3132 ttaattatatttcttctgtactgttaattgtgaatgttaagtaagaaagtccttct 3191  
2797 TTATTATTATTGTCTGTGAGCTTAATTGTGAATGTAAATGCAATTAAGTCCCTTGT 2856  
3192 agatgg 3197  
2857 AGATGG 2862

RESULT 3  
AX193124 2265 bp DNA PAT 15-AUG-2001  
LOCUS AX193124  
DEFINITION Sequence 691 from Patent WO0149716.

ACCESSION AX193124  
VERSION AX193124.1 GI:15211077  
KEYWORDS  
SOURCE  
ORGANISM human.  
Human sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2265)  
AUTHORS Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,  
Stolk,J.A., King,G.E., Wang,Y., and Jiang,Y.  
Compounds for immunotherapy and diagnosis of colon cancer and  
methods for their use  
Patent: WO 0149716-A 691 12-JUL-2001;  
CORIXA CORPORATION (US)  
FEATURES  
Location/Qualifiers  
source 1..2265  
/organism="Homo sapiens"  
BASE COUNT 679 a 484 c 522 g 580 t  
ORIGIN  
Query Match 32.1%; Score 1030; DB 6; Length 2265;  
Best Local Similarity 83.2%; Pred. No. 5.3e-252;  
Matches 2001; Conservative 0; Mismatches 260; Indels 145; Gaps 67;  
OY 801 ttgcctgtgttggaaggtatctactcaactcttcaatgtgaagaagaaggtat 860  
DB 2 TTGGCTGTGTTAAAGAGGATTAATCAACTCTTCTACATGTAAAGAAAGGATAT 61  
OY 861 tccccggaagatttcagtgacagatcacagaacatttgaccacgaagaagaacattcc 920  
DB 62 TCCCGG--GAAGATTTCAGTGTACACTATCAGAACTTTGACCCGAGAGAAACTTCC 120  
OY 921 atggcctatcaagactgtcatagtgaaatctacagctgtttaaagatgtatctggcaca 980  
DB 121 ATGGCTTATCAAGACTTGCATAGTGAATTTACTGCTGTGTTAAAGATGTAATTTGGCACA 180  
OY 981 tctgttcttgacagacgtatattcttactgtgaagacatctcgttaccacaagattctg 1040  
DB 181 TCTGTTATGACACACAGCTGTAATCTTACTGTAA--GCACATCTCTGTACCAAGA--TCTG 238  
OY 1041 aaatcgctgtatgataaacttltgttaactgttaacatagtaaacatttggcagaac 1100  
DB 239 AAATCGGTGC--TGATGACAAAG--TTTGTTAATGTAACTAATGTAACAAATTTGGCAGAAAC 296  
OY 1101 cacaagtacaatgagaagactgtgactgaggaataatataaagaacttataagtagc 1160  
DB 297 CACAAGTGACAATGAGAAGACTGTGACT--GAGAAATTAATTAAGCAA--TTAGAAGTAGC 354  
OY 1161 tcaagcaacttttcaaaacttgatgacacctgtcggtgtggtatgtatgaaggtggtg 1220  
DB 355 TCAACCAAC--TTTCTAAACTATGATTTGACCC-----TTGGGTGTGATTTATGCGCTG 407  
OY 1221 aaccaagactgctgtgagactgcctcaatgtgttgaatgacatgacatgtgcacaa 1280  
DB 408 TAACCAAGACTG--CGATGACTGCTCTCAAT--GCTTACATATGCCA--TTGCCAAATCTAC 462  
OY 1281 ctgcaagagcctaaccacagaagcccttctgtcgttgcctccagtcacagagtgctcga 1340  
DB 463 CTGCAAAAGGCTTAACCCACAGAGCCCTTCTGCGTTGCTTCCAGCTCA--AGTGCCTGA 521  
OY 1341 tgcctgacagcagaagcagaagcgaatgtcttaataaagaagtggtgtccctctga 1400  
DB 522 TGCCTGCAACGACACACAAAGC--AATGCTTAATAAAGAGGTGTGGGCCCTTGAGT 580  
OY 1401 gtgtgcgttgctgctccggtctacagaagaatgtcctaagtggaaactccaagaagtgc 1460  
DB 581 GTGCGTGGCTG-----CCGGGCTACACAGAGATGCTAATGGAACTCCCAAAGTGTGC 635  
OY 1461 attggactacaagtgactgtactgtlaagacaatctcaagctgactcactaattgt 1520  
DB 636 ATTT--GGCTACAGTGTGACTGTGTAAGGACAAATTTACACTGTATCTCACT--ATTGT 692

QY	1521	gggacacatgcgtgcatctgtctccttcacgcatgatgaatctgatgtatgtctacagaa	1580
Db	693	ggggcaccatccctggcattgctcattccacgaatgaatattgattgattgtcc-AGCA	751
QY	1581	gatacaataaacaagaagaaacalatttgaaacagaaacttgatctgacgaacattca	1640
Db	752	GATCAATAATACAAA--CGAAGCATATTGAAGA--GAGAACTTGATTGACGAAGACTTTCA	809
QY	1641	aaatctaaactgcgtctcgacagagcttcaaccaactctatgtagacataacgagcgtctc	1700
Db	810	AAATTTAAATCGGGGTTCG-ACAGGGCTTCAACCAATCT--TGGAGCGAAGAGGGAGCCTCTTT	867
QY	1701	cctcaagtcagatctaacggcctcccaagagacgacctagatgtaaaaaatcccgtagttca	1760
Db	868	CCTAAGGTCAGAGATACGGGCTCCAGAGA--CAGCCAGATGCAAAATCCC-----TATTCA	921
QY	1761	agaacagcagcatgcccccgccgctgactatagaatccatcagaatgtgnaaccgcca	1820
Db	922	AGACACACACACATG-CCCCGCCCTGACTATTAGAAAT--CATTAAGATGTGNAACCCGCCA	979
QY	1821	tggcccccaaccatattgacatctcattatctctagcaatgttttagaagaactgcacg	1880
Db	980	TGGCCCCAACCAATGACAGCTATTATT-----AAGTGTTTAGAAAGCTGATGG	1033
QY	1881	agaagtgcagccacacgtaaaagactctggccctccggagttctctccatctagacata	1940
Db	1034	AGAAATGAGCACCA--GTAAAGATCTGGCTCCGGGGTTTCTTCATCGATCGATGCTGC	1091
QY	1941	ctgcagatcccatctgcgaatgagcaagttgtgcattgtctctgcaaaagacatcaagct	2000
Db	1092	CAGCTCTCTGAATGGAAGTTGTGAATGTTGCA-----CGAATCCACT	1137
QY	2001	cactgtcctaaataagatcatctatgacatbaactatgtagctcatctatagcgtctgc	2060
Db	1138	CACTTGCT--AAATGAAGATATATGACTTA--AATGTATATATGCTATTATAGCGTTGTC	1194
QY	2061	tcaagagtggtgttctcttcaatcagtaaacaagtactgagacaatgcttatgggtttg	2120
Db	1195	AGAGA--GGTGGTTTTCTTCAATCAAGT--ACAAGATACGACACATAGTTA--GGGTGT	1249
QY	2121	ttctctaatcttctccctgtagggacaagaagcccaacttccaaactgagaaagc	2180
Db	1250	TTTTCTTAATTTT--CTGTGATGGGCACACAGAAC--ATTTCATATCTAGAGAAAG-	1305
QY	2181	ctccccaagcatgtctgtctccctctggccaacaacatgctctcttgatlaattgactaa	2240
Db	1306	CTCCCCACACATTTGCT---TGCTCTGGGGCAACATTTGCTTTGAGTTAAG--TGACTTA	1360
QY	2241	cttcccccggagacacataccgcataactcgtggaggtctccgggggagatgaaaggat	2300
Db	1361	TTCCCTTGGGAGACATA--CGCATCACTGTGAGAGTCCGAGGGG--ATGAGAAGGGAT	1415
QY	2301	accacaacatcttcatatgaggttcaacaactacacatctcgtgaaanaagtcagaaatagggaca	2360
Db	1416	ACCACACATCTTTCAAGGCTCACACAGCTCACTCTC--TGACAGTCAGAATAGGG--AC	1470
QY	2361	ctcgtctctatccctccaatgagagagatcttggccaacaaccccccttltttugaacca	2420
Db	1471	ACTGCTTTCTATCCCTCCAAATGGA--GAGATTCTGGCA-----GCTTTGA	1513
QY	2421	ggcccccaagagcttggaaactctgactccaaccaaaagaagacttggaaagagacatatct	2480
Db	1514	ACAGCCCAAGCTT--GCAACTTACCTC--ACCCAGAAAGACTTGGAA--GAAGCATATTC-T	1569
QY	2481	ttcagcttttccagagcgctgcttgggaatccagaaagcttlttgaigtctaaattgaa	2540
Db	1570	CTCAGCTTTTTCAGAGAGCGCTGCC--TGGGAATCCAGAAAC--TTTTTGATGCTAATTTGAA	1627
QY	2541	ggcctgagacataataatgtccatctatggttttaactacagcttlttgaacatgcta	2600
Db	1628	GGCTGTGACTAATAAATCTCAC---TAATGGGGTGCACCTCTCACTTTTGTAA--ATGCTA	1682

QY	2601	ggaagcagaagggccaaagagtaaaaaaacatgaaccgttaagaaagaaagagcaca	2660
Db	1683	GGAGCGAAGAGGGCGAGAGC---TAAAAACATGACCTGGTAGAAGAGAGAGCAAA	1739
QY	2661	ggaacatgagctggggagagatacaatltagagagagagcacttgatgcacactcgtctcc	2720
Db	1740	GGAAACGTGGGtGGGAGAGATCAATTAGAGAGGAGGACCTGGGATTCACCTTC-TTCCTT	1798
QY	2721	aggtccctccctccatgacagcaaaagacactctctaaatgaatgacctccgaagactg	2780
Db	1799	AGGTCGCCCTCCTCCAT-CACCAAGGAGCACTCTCTTAA-TCATGCCCTCCGGAAGACTG	1856
QY	2781	gctggagaaaggtttaaaaaaacaataaccagagtaagagaccttaggtgacagtcttg	2840
Db	1857	GCTGGAGAGAGGTTT-AAAAACAAAAATTCAGAGACT-AAGAGCCTTAGG---TCAGTTT	1911
QY	2841	aaaatltgagacaaactgtctgtgcacaaagtgccaaagcgcagactgtgtgtcaca	2900
Db	1912	GAATTTTGAGACAAACT--GTCGTGCAAAAGGGTGC---GAGAGGAGAGCTTGGTGCACGA	1966
QY	2901	gtcccaagcgtccaagcctctggagtgtaaggtctcttgagtggtgtgcaatggggcctcagc	2960
Db	1967	GT-CCAGCGCGTCAGCCTCGGGGGTGT-AGGTTTCTAGGCTGGCCATTGGGGCCTCACCC	2024
QY	2961	tctctgtgtgccccgaagcctcagctgtgtgcccacacacacacacacacacacaca	3020
Db	2025	TTTCTGTGTGACA-GAGGCTCACTGTGGCGACACACACACACACACACACACAA	2083
QY	3021	cacacacaaatgggggcaaccacacatccacatgtaaccaagcttlaacacaaatgltatag	3080
Db	2084	CACACACAAATGGGGGCAACACATCCAGTCAAGC-----TTTTACAAATGTATTAG	2137
QY	3081	tgctcccttatattcctaagccgcgtcctctaagtaatttatgtttatata	3140
Db	2138	TGTCCCTTTTATTCTTAAT-GCCTGTGCTCTTAAAGATTATTTTATTGTATTATTA	2196
QY	3141	tctgtctgactgtaattgtaatgtaatgtaatgcaataagtgacctgtgtagatgaaa	3200
Db	2197	TTTGTCTGTGACTTAAATTGTGAATGTATGTATGTCAATTAAGTGCTTTGTTAAAAAAA	2256
QY	3201	aaaaa 3206	
Db	2257	AAAAA 2262	

RESULT	4
AK000070	AK000070 2265 bp mRNA PRI 22-FEB-2000
LOCUS	
DEFINITION	Homo sapiens CDNA FLJ20063 f1s, clone COL01524.
ACCESSION	AK000070
VERSION	AK000070.1 GI:7019918
KEYWORDS	Oligo capping; f1s (full insert sequence)
SOURCE	Homo sapiens colon CDNA to mRNA, clone:lib:COL clone:COL01524.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (sites)
AUTHORS	Kawabata,A., Hikiji,T., Kobatake,N., Inagaki,H., Ikema,Y., Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Odayashi,M., Nishii,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
	NEDO human cDNA sequencing project
	Unpublished (2000)
TITLE	2 (bases 1 to 2265)
JOURNAL	Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.
REFERENCE	Direct Submission
AUTHORS	Submitted (15-FEB-2000) to the DDBJ/EMBL/GenBank databases. Sumitomo Sugano, Institute of Medical Science, University of Tokyo, Dep'tment of Virology; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdm@iims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
TITLE	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert
JOURNAL	
COMMENT	

sequencing: Research Association for Biotechnology; cDNA library construction, 5'-6' 3'-end one pass sequencing; Department of Virology and Human Gene Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES  
source Location/Qualifiers

1. .2265  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COL01524"  
/clone\_1db="COL"  
/issue\_type="colom"  
/note="Cloning vector pME18SFL3"  
121. .954  
/note="unamed protein product"  
/codon\_start=1  
/protein\_id="BAA90925.1"  
/db\_xref="gi:7019919"  
/translation="MAYODLSEITSLFKDVGTSVYGQVILFVTSLSRSEMRAD  
DKFVNVITLITAEITSDNEKTEIKKAIKATSSSNFNLDTLRDQYGCNQTAD  
CLNLACDCKSDLRPNPSPCVASLICKGADCAOAKOOLIKKSGAPACACVPCY  
OEDANGNOKCAFISGLDCKDFOLIITIVGTAGVILSMITLITATSSNNKTKH  
IEENLIDEDFQNKLNKSTGFNLGAEESVPEPKVITASRDSOMNPYSRHSMPRP  
Y"

CDS

BASE COUNT 679 a 484 c 522 g 580 t  
ORIGIN

Query Match 32.1%; Score 1030; DB 9; Length 2265;  
Best Local Similarity 83.2%; Pred. No. 5.3e-25;  
Matches 2001; Conservative 0; Mismatches 260; Indels 145; Gaps 67;

OY 801 ttgttcgtgtttagaagggtatctactacaactctctcacaagttaagaagaaggtat 860  
DB 2 ttgttcgtgtttagaagggtatctactacaactctctcacaagttaagaagaaggtat 61  
OY 861 tccctcgagaagattcctagtgacatcatcaagaacattgaccacgaagaagaacattcc 920  
DB 62 tccctcgagaagattcctagtgacatcatcaagaacattgaccacgaagaagaacattcc 120  
OY 921 atgacctatcaagacttgcatatgagaaatctactagcttgattaaagatgattggcaca 980  
DB 121 atgacctatcaagacttgcatatgagaaatctactagcttgattaaagatgattggcaca 180  
OY 981 tctgtttatgagacactgtatcttctctgtaaggcacatctctgttaccagaattctcg 1040  
DB 181 tctgtttatgagacactgtatcttctctgtaaggcacatctctgttaccagaattctcg 238  
OY 1041 aaatcgctgtatgacaaagttctgttaattgtaacaatagtaacaatttggcagaac 1100  
DB 239 aaatcgctgtatgacaaagttctgttaattgtaacaatagtaacaatttggcagaac 296  
OY 1101 cacaagtgaacaatgagaagacttgatgagaaataataaagaacatttataagtagac 1160  
DB 297 cacaagtgaacaatgagaagacttgatgagaaataataaagaacatttataagtagac 354  
OY 1161 tcaagaacatttcaaaactatgtagcctgcgcgtggtgattgattgtagggctggg 1220  
DB 355 tcaagaacatttcaaaactatgtagcctgcgcgtggtgattgattgtagggctggg 407  
OY 1221 aaccagaagctgctgtagtactcaatggttttagatcagatgtagtcaatgctgac 1280  
DB 408 taaacagactg--cggatgactgctcaat--ggtttagatcagatgtagtcaatgctgac 462  
OY 1281 ctgcaaaagcttaaccacagaagccttctgctgcttgcctcagatctcaagatgcttca 1340  
DB 463 ctgcaaaagcttaaccacagaagccttctgctgcttgcctcagatctcaagatgcttca 521  
OY 1341 tgcctgcaacgacagcaagaacgaatcttaataaagaagaatgtagggctcccttca 1400  
DB 522 tgcctgcaacgacagcaagaacgaatcttaataaagaagaatgtagggctcccttca 580  
OY 1401 gtgttcgtgtgctgctgctgcttaccagaagaatgtagtcaatggaactgcacaaagtgtgc 1460

DB 581 gtagcgtggcgtg-----cccggtaccagagaagatcttatgggaactgcacaaatgctgc 635  
OY 1461 atttggtctaacgtgtagctgactgactgtaagagacaattcgaactgactcctactattgt 1520  
DB 636 atttt--ggctacagtgactgcactgtaagagacaatttgcactgactcctactattgt 692  
OY 1521 gggacacacgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1580  
DB 693 gggacacacgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 751  
OY 1581 gatcaataaacaagaagaac 1640  
DB 752 gatcaataaacaagaagaac 809  
OY 1641 aatctaaactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 1700  
DB 810 aatctaaactgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 867  
OY 1701 cctcagtgctgagatgtagcgcctcgaagacacgctgtagatgtaaaatccgtagttca 1760  
DB 868 cctcagtgctgagatgtagcgcctcgaagacacgctgtagatgtaaaatccgtagttca 921  
OY 1761 aagacacagacatgccccgc 1820  
DB 922 aagacacagacatgccccgc 979  
OY 1821 tggcccccaacacatgtaacatctatctatctagcagtggtttagaagaagactgacg 1880  
DB 980 tggcccccaacacatgtaacatctatctatctagcagtggtttagaagaagactgacg 1033  
OY 1881 agaagtggac 1940  
DB 1034 agaagtggac 1091  
OY 1941 ctgcagtcctcctgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 2000  
DB 1092 ctgcagtcctcctgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatgca 1137  
OY 2001 caactgtcaaaataagaatctatgacatgaatgacatgacatgacatgacatgacatgac 2060  
DB 1138 caactgtcaaaataagaatctatgacatgaatgacatgacatgacatgacatgacatgac 1194  
OY 2061 tcaagagtggtggttcttctcaatgaacacacacacacacacacacacacacacacacac 2120  
DB 1195 agaga-----gctggttttcttcaatcagc--acaaactgacacacacacacacacacac 1249  
OY 2121 ttcttcaatcttctcctctgtaggcaagaacacacacacacacacacacacacacacacacac 2180  
DB 1250 ttcttcaatcttctcctctgtaggcaagaacacacacacacacacacacacacacacacacac 1305  
OY 2181 ctccccacatgcttgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2240  
DB 1306 ctccccacatgcttgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1360  
OY 2241 ctccccacatgcttgcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2300  
DB 1361 ttccccctggagacacatp---cgcattcaactgtagaggtccgagggg--atgagaggggat 1415  
OY 2301 acccaccatcttcaatgtaggcaagaacacacacacacacacacacacacacacacacacac 2360  
DB 1416 acccaccatcttcaatgtaggcaagaacacacacacacacacacacacacacacacacacac 1470  
OY 2361 cctgctctatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2420  
DB 1471 actgctctatcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1513  
OY 2421 ggcctccagagcttggaac 2480  
DB 1514 acagccccaagctt--gcacctagcctc--acccaagaagaactggaac--gagacacatc--t 1569  
OY 2481 ttcagcttcttcaagaagcgtgcttgggaatcaggaagcttcttgaatgtaataa 2540

Db 1570 CTCAGCTTTTCAGAGGCGCTGCC-TGGGAATCCAGGAC-TTTTTCATGCTAATTAGAA 1627  
 Qy 2541 ggcctgactataaataatgcatatgaggttttaactacagttttgaacatgcta 2600  
 Db 1628 GGCCTGGAGTAAATATGTCAC-----TATGGGGTGCACCTTACAGTTTGTGAA-ATGCTA 1682  
 Qy 2601 ggaagcagaagggccagagagtaaaaaaacatgacctgtgtagaaggaagagggcaaa 2660  
 Db 1683 GGAGGCGAAGGCGCAGAGAG---TAAACCAATGACCTGTGAAGAGAAAGAGGCGCAAA 1739  
 Qy 2661 ggaacatgggtgggagagatcatagagagagagacactgtgagatccactgttccct 2720  
 Db 1740 GGAACATGGGTGGGAGAGATTAATAGAGAGGAGGACCTGGATCCACTTC--TTCCCTT 1798  
 Qy 2721 agtccctcctcctcatgcaagcaagagacacttctctaagtcagtcctcccgaaagctg 2780  
 Db 1799 AGTCCCTCCTCCAT-CAGCAAGAGGACACTTCTTAA-TCATGCCCTCCCGAAGACTG 1856  
 Qy 2781 gctggagagaggtttaaaaaaacaataaccagagagtaagaagccttagagtcagtttg 2840  
 Db 1857 GCTGGAGAGAGGTTT-AAAAACAAATAATCCAGAGT-AMAGGCTTAGG---TCAGTTT 1911  
 Qy 2841 aaattggaagaacaactgtctgtgcaaaagggtgcaagagcggagcctgtgtgtctcaga 2900  
 Db 1912 GAATTTGGAGACAAACT--GTCTGGCAAAAGGGTGC---GAGAGGAGCCTTGTGCTCAGGA 1966  
 Qy 2901 gtcccaagccgtccagcctcggggtgttaaggtctgtgagtggtgcatagggggacctcagcc 2960  
 Db 1967 GT-CCAGCGCTCAGCCCTGGGGGTGT-AGTGTCTGAGGTGTGCCATTTGGGGCCTCAGCC 2024  
 Qy 2961 ttctctgtgacccgaggtcagctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3020  
 Db 2025 TTCTCTGTGTGCA-GAGGCTGAGCTGTGGCCACACACACACACACACACACACACACACA 2083  
 Qy 3021 cacacaataatgggggagagac 3080  
 Db 2084 CACACACAAATGGGGGACACACACACATCCAGTACAAGC-----TTTTACAAATGTATTAG 2137  
 Qy 3081 tgcctcttatttatttataatgacctgtcctcttaaaagtatttatttatttatttatttatt 3140  
 Db 2138 TGTCCTTTTATTTCTTAAT-GCCTTGCTCCTTAAAGTTATTTATTTATTTATTTATTA 2196  
 Qy 3141 ttgtcttgcactgttaattgttgatgtgaatgtaataaagtgcttctgttgaatgtgaaa 3200  
 Db 2197 TTTGTCTGTGACTGTATATGTGAATGTAATGCAATTAAGTGCCTTTGTTAAAAAAA 2256  
 Qy 3201 aaaaaa 3206  
 Db 2257 AAAAAA 2262  
 RESULT 5  
 AX011684 2263 bp DNA PAT 06-SEP-2000  
 LOCUS AX011684 Sequence 80 from Patent WO955858.  
 DEFINITION AX011684  
 ACCESSION AX011684  
 VERSION AX011684.1 GI:9998208  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and  
 Pilsarsky,C.  
 TITLE Human nucleic acid sequences obtained from pancreas tumor tissue  
 JOURNAL Patent: WO 955858-A 80 04-NOV-1999;  
 SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DAHL EDGAR (DE); HINZMANN  
 BREND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN  
 (DE); PILARSKY CHRISTIAN (DE)  
 FEATURES  
 source 1..2263  
 /organism="Homo sapiens"

BASE COUNT 689 a 483 c 520 g 571 t  
 ORIGIN  
 Query Match 31.5%; Score 1009; DB 6; Length 2263;  
 Best Local Similarity 83.0%; Pred. No. 1,2e-246;  
 Matches 1980; Conservative 0; Mismatches 260; Indels 145; Gaps 67;  
 Qy 822 attactacaactcttctcatatgtaagaagaagagatctccctggaagatctcagt 881  
 Db 1 ATTACGACACTCTTCTACATGTAAGAAGAAAGTATTCCTCG-GAAGATTTCAGTG 59  
 Qy 882 acagatcagaacaatttgaccagaagaagaacaattccatgacctatcaagattgcat 941  
 Db 60 ACAGTATCAGAAACATTGTGACCAGAGAAACATTCCATGCGCTATCAAGACTTCAT 119  
 Qy 942 agtgaattactagctgttttaagaatgattttgcaacatctgtttatgacagacgta 1001  
 Db 120 AGTGAATTAATTAAGATGTAATTTGGCACATCTGTTATGACACACTGTA 179  
 Qy 1002 attcttactgtaaggaacatctctgtcaccagaattctgaatgctgtctgtatgacaag 1061  
 Db 180 ATTCTTACTGTA-GCACATCTCTGTCAACCAAG-TCTGAATTCGTGC-TGATGACAAG 236  
 Qy 1062 ttctgttaatgtaacaatgtaacaatttggcagaacaacagtgacaatgagaagac 1121  
 Db 237 -TTTGTAAATGTAACAATAGTAACATTTTGGCAAGAACCAACAAGTGACATGAGAAGAC 295  
 Qy 1122 tgtgactggagaataatgaataaacaattataagttgctcaagcaactttctaaacta 1181  
 Db 296 TGTGACT-GAGAAAAATTATTAAGCA--TTAGAATGATGCTCAAGCAAC-TTCTTAACCTA 352  
 Qy 1182 tgaatgacccctgtcgtgtgtgattgattgaaggctggaacaaagactgctgtagtac 1241  
 Db 353 TGATTGACCCCTTC-----GGTGTGATTATTATAGCTGTAAACCAAGCTGC--GGATGAC 404  
 Qy 1242 tgcctcaatgggttttagcatgcatgtgcaaatgtgcaactgtcaaaaggcctaaccacag 1301  
 Db 405 TGCCCTCAAT-GGTTTACATGCGATTCG--AAATCTGACCTGCAAAAGCCCTAACCCACAG 461  
 Qy 1302 agcccttctgcgttgccttcaagctcaaggtgctgagtcggaagcaagaacaaacaa 1361  
 Db 462 AGCCCTTCTGCGTGTCTTCAGTCTCA-AGTGTCTGATGCTCGACACGACGCAAA 520  
 Qy 1362 gcaaatgcttaataaagaagagtggtgggtccctcgaagtgtgtgcgttgcgtgcgcgt 1421  
 Db 521 GC-AATGCTTAATTAAGAAGAGTGTGGGCCCTGTAGTGTGCTGCGTGC-A 574  
 Qy 1422 ctaccaggaagatgtctaaatgggaactgtccaaagtgcatattgggtctacagtggactg 1481  
 Db 575 CTACACAGGAAGATCTAATGGAAGTCCCAAAAGTGTGCAATTT-GCTTACAGTGAATCGC 633  
 Qy 1482 acgtgaaggaacaattcaactgcatgctcctactatattgttggaacacacgctgcatgt 1541  
 Db 634 ACTGTAAGACAAATTTTACGTGATCCTCACT-ATTGTGGGACCAATCCCTGGCATTTG 691  
 Qy 1542 catctcagcatgataatgtcatgtatgttcaactgacagatcaataaacaagaagcgaag 1601  
 Db 692 CATTCCTCAGCATGATTAATTCATTTGATTGTTCAC-AGCAAAATCAAAATACAAA-CCAG 749  
 Qy 1602 cataatgaagaacgagaactgtatgacgaagacttcaaaatctaaactgagtgca 1661  
 Db 750 CATATTGAAGAA-GAGAACTTGATTGACGAAGACTTTCAAAATCTAATAACTGCGGTGC-A 807  
 Qy 1662 caggtctcaccatctatgtagatcaacgagaggtcttccctcaggtcgaagatgaagcc 1721  
 Db 808 CAGCTTACCAAACT-TGAGACGAAGAGGAGGCTCTTTCTTAAGTCAAGATTAACGGCC 866  
 Qy 1722 tccaagagacgcgtatgatacaaaaatccgtagttttcaagacacagcatgacgcccgcg 1781  
 Db 867 TCCA-GAGACAGCCAGATGCAAAATCCC-----TATTCAAGACACAGCAGCATG-CCCCG 919



QY	1782	gctgacactatagaatccatccagaatggtgaacccgcctatgcccccaacatatgtca	1841
Db	920	CCCTGACCTATTAGAAAT-CATTAAGAAATGTGAACCCGCCATGGCCCCCAACCAATGTACAA	978
QY	1842	tatcatattcttagcaggtgttttagcagaagctgcatagtgagagtgcacacagtaag	1901
Db	979	GCTATTATT-----AGAGTGTTTAGAAAAGACTGATGGAGAAAGTAGAGACCA--GTAA	1030
QY	1902	acttcgctccggyagtttcttcttcacatagacatctcagctccatctgcgat	1961
Db	1031	GATCTGGCCTCCGGGGTTTTTCTTCATCTGCATCTGCACACCTCTCTGAAATGSAATTT	1090
QY	1962	ggcaagctgtgcaatgctcttcgcaacgacatccacgcctactgtctaaataagaatct	2021
Db	1091	GTGAAATTTTGGCAACG-----AATCCAGCTCACTTGGCT-AAATTAAGAAATCT	1135
QY	2022	atgacataacatgtagctcgatgcctattagagcgtgtgctgagagtggtttcttc	2081
Db	1136	ATGACATTTA--AATGTATGATGATGATTAAGCCCTGTGTAGAGA--GGTGGTTTTCTTC	1190
QY	2082	aatacgaacaagaatctgagacaatgcttaaggggtttgttcttaattcttcccg	2141
Db	1191	AATACAGT-ACAAAGTACTGAGAACAAATGGTTA-GGGTGTGTTTCTTAATTTCTTT-CCGCG	1247
QY	2142	tagggcaacaagaccocatctccaaatctagagaaagccctcccaagatgtctgtc	2201
Db	1248	TAGGGCAACAAGAAC--ATTTCATCTTAAGAGAAAG-CTCCCCAGACTTGGCT---TG	1300
QY	2202	cctgtggccaacatgctctctttagttaaagttagcttaacttcccctggaagcaatcc	2261
Db	1301	CTCCTGGGCAAACTGCTCTTTAAGTTAAG-ATGACTTAATTTCCCTGGAGACATA---C	1356
QY	2262	gcatacaactgtgaggttcgaggggagtagaagaaggaataccacaacttcaaggt	2321
Db	1357	GCATCACTGTGGAGGTCCGAGGG--ATGAGAAGGAAATACCCACACTTTCAAGGTC	1414
QY	2322	cacaagctacactctcgtgacaagtaagaataggggacactgtcttataccctcaatg	2381
Db	1415	ACAAGCTCACTCTC--TGACAGTCCAGATAGAGG--ACACTGCTTATCCCTCAATG	1469
QY	2382	gagagattcttgccaaccccccttttttgaaaacagagcccccaagacttggcaac	2441
Db	1470	GA-GAGATTTGSCAA-----CCTTGTACAGGCCCAAGACTT-GCAACC	1511
QY	2442	tagcctaacccaagaagacttgaagaaggacatatcttcaagctttcagaagaggt	2501
Db	1512	TAGCCTC-ACCCAAAGAAGATGGAAA-GAGACATATC-TCTAGCTTTTTCAGAGGGCT	1568
QY	2502	gccttgggaatccaagaagcttlttgatgctaattagaagccttgcataataatgtc	2561
Db	1569	GCC-TGGGAATCCAGGAAAC-TTTTATGCTAATTATGAAGGCGCTGACATAAAATGTCA	1626
QY	2562	catctatgggttttaattcaacagtttttgaacatgtctagagagcagaacgggccaag	2621
Db	1627	CTAT---GGGGTGCACTCTACAGTTTTTGAA-ATGCTTGAGGAGCAAGAGGGCGAGAG	1681
QY	2622	agtaaaaaaacatgacctgttgaagaagagagagcaagaagaacttgggttgggaagctc	2681
Db	1682	---TAAAAAACATGACTGTGTAAAGAGAGAGAGCAAGCAAACTGGTGGGAGAGATC	1738
QY	2682	aattagagaggaagcaacttggatccaactcgttcccttaagtctccctccatcgtcgc	2741
Db	1739	AATTAGAGAGAGGACCTGGATGCCACCTTC-TTCCTTAGTCCCTCTCTCCAT-CAGC	1796
QY	2742	aaaggagcacttctctaagtcacatgcccctccgaagactgtcgtgggaagaagttaaaaa	2801
Db	1797	AAAGAGCACTTCTTAA-TCATGCCCTCCCGAAGACTGTGGGAGAGAGTTT-AAAAA	1854
QY	2802	caaaaaatccagagtaagaagagccttaggtcagttttttaaatttggagacaactgtc	2861
Db	1855	CAAAAAATCCAGAGAT-AGAGGCTTTAGG---TCACTTTGAATTGGAGACAAACT-GT	1908
QY	2862	tttgcaaaaggttcccaagagcggagctgttgcataagagttccacagccgtccaagctcgg	2921

[illegible]



QY	2412	tgaaacacaggcccccagagcttggcaacctaagcttaacccaagaagactgtgaagag	2471
Dd	153672	--AACCTTTGAAACAGCCACGAGCTGCAACCTAGGCTC-ACCCAAAGAAGACTGGAAA-GAG	153617
QY	2472	acatactcttcaacttcttcaaggagcgctgcttggaaatccagaagcttcttgatgc	2531
Dd	153616	ACATATC-TCTCAGCTTTTTCAGGAGGCGTGC-TGGGAAATCCAGAAC-TTTTGTATGC	153560
QY	2512	taattagaagcccttggacataataatgtccatctatctatctgtttaaattcaagttttg	2591
Dd	153559	TAAATGAAAGGCGCTGACATAAAATGTCACAT-ATGGGGGTGCACCTACAGTTTGTG	153504
QY	2592	aactatgctaaaggcagaacgagccagagatataaaaaacatgacttgttaagaagag	2651
Dd	153503	AA-ATGCTAGGAGGAGAGAGGGGAGAGAGT-AAAAAACATCACTGGTATAGAGAG	153448
QY	2652	agaagcaagaagaacttggcttggagagatcaattagagaagaagcacttggatccact	2711
Dd	153447	AGAGCAAAAGCAACTGGGTGGGAGGATCAATTAGAGAGAGGACCTTGGATCCACT	153388
QY	2712	tcgttccctttagtccctccctcccatctgcagcaagaagagcacttcttaagcatgccc	2771
Dd	153387	TC-TTCCTTAGGTCCCTCCTCCAT-CAGCAAAAGAGACACTTCTTAA-TCATGCCCTCC	153321
QY	2772	cgaagactgcgttggagagaagtttaaaaaaacaataccaagaatgaagaccttaagg	2831
Dd	153330	CGAAACACGCGCTGGGAGAGGTTT-AAAAACAAAAATCCAGGAGT-MAAGGCTTAGG-	153272
QY	2832	tcaagtttgaatatgtgagacaacttgccttgcgaaggtgtccgaagcggagcttgc	2891
Dd	153273	--TCAGTTTGAATTTGGAGACAAACT-GTCTGGCAAAAGGTGC--GAGAGGAGCTTG	153221
QY	2892	tgctcaagagtcaccagccctcagcctcgggtgttaaggtctctgagtgtgcacatggg	2951
Dd	153220	TGCTCAGGAGT-CCAGCGCGTCCACCTCGGGGTGT-AGTTTCTGAGGTGTGCATTGGG	153163
QY	2952	gctcagccctctctgtgtgaccgaggtcagctgtgtggccacaacacacacacacac	3011
Dd	153162	GCTCAGGCTTCTCTGGTGA-CAGAGGCTCAGCTGTGGCCACCAACACACACACAC	153104
QY	3012	acacacacacacacacaaatggggagcaaccacaatccacgtlaaccaagctttaacca	3071
Dd	153103	ACACACACACACACCAATGGGGCAACCAATCCAGTCAACAGC-----TTTTACAAA	153050
QY	3072	tgtaatagtgcccttcttaattctaaatagccctgcctcttaaaagttaattatgt	3131
Dd	153049	TGTTATGTGTGCTTTTATTCTAAT-GCCTGTGCTCTTAAAGTTATTATTATTTG	152991
QY	3132	ttatataatttcttctgactgttaattgtgaatgtaatgcaataaagtgccctgtc	3191
Dd	152990	TTATATTATTGTCTTCATGCTTAATTGTAATGTGTAATGCAATTAAGTGGCTTTGTT	152931
QY	3192	agatgg 3197	
Dd	152930	AGATGG 152925	
RESULT	7		
LOCUS	AX079657	536 bp DNA	PAT 22-FEB-2001
DEFINITION	Sequence 401 from Patent WO0107611.		
ACCESSION	AX079657		
VERSION	AX079657.1	GI:13159226	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.		
TITLE	Baker, K.P., Goddard, A. and Wood, N.I.		
JOURNAL	Human polypeptides and methods for the use thereof		
	Patent: WO 0107611-A 401 01-FEB-2001;		
	Genentech, Inc. (US)		

FEATURES	Location/Qualifiers
SOURCE	1. .536 /organism="Homo sapiens" /db_xref="taxon:9606"
BASE COUNT	166 a 156 c 77 g 134 t 3 others
ORIGIN	
Query Match	11.4%; Score 365; DB 6; Length 536;
Best Local Similarity	90.5%; Pred. No. 3e-12;
Matches 503; Conservative	0; Mismatches 33; Indels 20; Gaps 10;
QY	279 taacaaccaagaactcgcgactagtggtctctacagtagtcgacgtctataccactga 338
DB	1 TTAACACACACAAACTG-GANTAGTGGTCCACAGTGTGACCTGACATACCATGAA 59
QY	339 ctaatttcgccgaacgcgtctagacacagaaataacccctctctccaaacact 398
DB	60 CTAAATTT-CCCTGAACGTGTAGCACACAGAAATACACNTCTTCTCCACACAGTACT 118
QY	399 tcaactgtcccccacataatagtaacacatagttctctccacaattccacactgtcc 458
DB	119 TCACCTGCTCCCCCATATATTAG-ACACATGTTCTCTCCACATTCCTACACCTGCTCC 177
QY	459 ccataatagtaacacatagttctctccacaattccataactcgtctgcagacgtag 518
DB	178 CCCATTAATTAGTACACATAGTTCCTCCACAAATTCCTTAACACACTGTGACACAGTGG 237
QY	519 tcaaccacaatagttaattcagttgcctctgaacctctgaataaccggttaactccaa 578
DB	238 TCACCCACAAATGTAAATTTCA-TTAGCTACCTGTACATAATCACCGCTTCATCTCCAAA 296
QY	579 tgatgattaatcacaatggtctctctgaacaacaagtaacaatgaatgtcccca 638
DB	297 TGATGATTAA-TCACATAGTTCCTCTCTGAAACACAAAGTACATGAATGATGCCCA 355
QY	639 ccacagaagaacaatcaatctcagtggtccctccactggcaccgcttatttgatgacc 698
DB	356 CCACAGAAGACATCAATCATCATCAG-GGCTCCCATCGGACGCGTTTATGGAGCACAGC 414
QY	699 atgcacgcctaaacagcagcagtggtcccaagatctctgccaagatgatccccgtgca 758
DB	415 A-----CCCTTAACAGCACAG-GTCCCGACCAATCTTGCCA--AGATGATCCCTGTGCA 465
QY	759 gataattcgtatctgtgttgtaagcttgcatacaaaagtttggcgtgtttagaag 818
DB	466 GATTAATTCCTTAATGTGCTTAAGCTGC-----ATAATACAAAGTTTGGCCGTGTTAGAAG 520
QY	819 ggtattactacaactc 834
DB	521 GGTATTACTACCACTC 536
RESULT 8	
LOCUS	AX192560/c
DEFINITION	Sequence 127 from Patent WO0149716.
ACCESSION	AX192560
VERSION	AX192560.1 GI:15210523
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 417) Xu,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J., Stolk,J.A., King,G.E., Wang,T. and Jiang,Y. Compounds for immunotherapy and diagnosis of colon cancer and methods for their use Patent: WO 0149716-A 127 12-JUL-2001.
JOURNAL	CORIXA CORPORATION (US)
FEATURES	Location/Qualifiers
SOURCE	1. .417



Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: L8653  
Center Clone name: 816\_1\_3

NOTE: This record contains 141 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
758 857: contig of 757 bp in length  
758 857: gap of 100 bp  
858 1584: contig of 727 bp in length  
1585 1684: gap of 100 bp  
1685 2449: contig of 765 bp in length  
2450 2549: gap of 100 bp  
2550 3306: contig of 757 bp in length  
3307 3406: gap of 100 bp  
3407 4163: contig of 757 bp in length  
4164 4263: gap of 100 bp  
4264 5024: contig of 761 bp in length  
5025 5124: gap of 100 bp  
5125 5878: contig of 754 bp in length  
5879 5978: gap of 100 bp  
5979 6734: contig of 756 bp in length  
6735 6834: gap of 100 bp  
6835 7583: contig of 749 bp in length  
7584 7683: gap of 100 bp  
7684 8446: contig of 763 bp in length  
8447 8546: gap of 100 bp  
8547 9304: contig of 758 bp in length  
9305 9404: gap of 100 bp  
9405 10161: contig of 757 bp in length  
10162 10261: gap of 100 bp  
10262 11003: contig of 742 bp in length  
11004 11103: gap of 100 bp  
11104 11831: contig of 728 bp in length  
11832 11931: gap of 100 bp  
11932 12694: contig of 763 bp in length  
12695 12794: gap of 100 bp  
12795 13547: contig of 753 bp in length  
13548 13647: gap of 100 bp  
13648 14410: contig of 763 bp in length  
14411 14510: gap of 100 bp  
14511 15268: contig of 758 bp in length  
15269 15368: gap of 100 bp  
15369 16119: contig of 751 bp in length  
16120 16219: gap of 100 bp  
16220 16986: contig of 767 bp in length  
16987 17086: gap of 100 bp  
17087 17816: contig of 730 bp in length  
17817 17916: gap of 100 bp  
17917 18663: contig of 747 bp in length  
18664 18763: gap of 100 bp  
18764 19505: contig of 742 bp in length  
19506 19605: gap of 100 bp  
19606 20362: contig of 757 bp in length  
20363 20462: gap of 100 bp  
20463 21180: contig of 718 bp in length  
21181 21280: gap of 100 bp  
21281 22041: contig of 761 bp in length  
22042 22141: gap of 100 bp  
22142 22903: contig of 762 bp in length  
22904 23003: gap of 100 bp  
23004 23764: contig of 761 bp in length  
23765 23864: gap of 100 bp

23865 24598: contig of 734 bp in length  
24599 24698: gap of 100 bp  
24699 25463: contig of 771 bp in length  
25470 25569: gap of 100 bp  
25570 26322: contig of 753 bp in length  
26323 26422: gap of 100 bp  
26423 27173: contig of 751 bp in length  
27174 27273: gap of 100 bp  
27274 28034: contig of 761 bp in length  
28035 28134: gap of 100 bp  
28135 28876: contig of 742 bp in length  
28877 28976: gap of 100 bp  
28977 29730: contig of 754 bp in length  
29731 29830: gap of 100 bp  
29831 30587: contig of 757 bp in length  
30588 30687: gap of 100 bp  
30688 31437: contig of 750 bp in length  
31438 31537: gap of 100 bp  
31538 32297: contig of 760 bp in length  
32298 32397: gap of 100 bp  
32398 33152: contig of 755 bp in length  
33153 33252: gap of 100 bp  
33253 34009: contig of 757 bp in length  
34010 34109: gap of 100 bp  
34110 34856: contig of 747 bp in length  
34857 34956: gap of 100 bp  
34957 35714: contig of 758 bp in length  
35715 35814: gap of 100 bp  
35815 36564: contig of 750 bp in length  
36565 36664: gap of 100 bp  
36665 37401: contig of 737 bp in length  
37402 37501: gap of 100 bp  
37502 38273: contig of 772 bp in length  
38274 38373: gap of 100 bp  
38374 39132: contig of 759 bp in length  
39133 39232: gap of 100 bp  
39233 39996: contig of 764 bp in length  
39997 40096: gap of 100 bp  
40097 40858: contig of 762 bp in length  
40859 40958: gap of 100 bp  
40959 41726: contig of 768 bp in length  
41727 41826: gap of 100 bp  
41827 42591: contig of 765 bp in length  
42592 42691: gap of 100 bp  
42692 43453: contig of 762 bp in length  
43454 43553: gap of 100 bp  
43554 44307: contig of 754 bp in length  
44308 44407: gap of 100 bp  
44408 45157: contig of 750 bp in length  
45158 45257: gap of 100 bp  
45258 45977: contig of 720 bp in length  
45978 46077: gap of 100 bp  
46078 46841: contig of 764 bp in length  
46842 46941: gap of 100 bp  
46942 47693: contig of 752 bp in length  
47694 47793: gap of 100 bp  
47794 48538: contig of 746 bp in length  
48539 48639: gap of 100 bp  
48640 49387: contig of 748 bp in length  
49388 49487: gap of 100 bp  
49488 50243: contig of 756 bp in length  
50244 50343: gap of 100 bp  
50344 51101: contig of 758 bp in length  
51102 51201: gap of 100 bp  
51202 51960: contig of 759 bp in length  
51961 52060: gap of 100 bp  
52061 52818: contig of 758 bp in length  
52819 52918: gap of 100 bp  
52919 53662: contig of 744 bp in length  
53663 53762: gap of 100 bp  
53763 54516: contig of 754 bp in length  
54517 54616: gap of 100 bp  
54617 55380: contig of 764 bp in length



Db	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	ORGANISM	SOURCE	TITLE	COMMENT
OY	1203	gattatgtagggcgctggggaacccaagacggtgcgtgagctgcctcaatgaaggttttagctg				human.			
OY	1203	gattatgtagggcgctggggaacccaagacggtgcgtgagctgcctcaatgaaggttttagctg				human.			
Db	1108	GTGTATTATTATTATTCGCTGTGTGAAGAGTGGCAGCAATTTTTCGGGAATGGTTTTCACATG				human.			
OY	1263	cgatgtgcaaatgtcgacactgtgcaagagcgtcaaacacacagagcccttcgtgctgtcc				human.			
Db	1168	C-ACGTGCAGAAACCTGGGCTGGAGAGACTGAAACCCGAGTTTCCTTTTGTGTTGCT---				human.			
OY	1323	agctcagaagtgtcctgtatgtcgtcaacgcacagcacaagc--gaatgcttaataaaga				human.			
Db	1223	---CCAAACATGTTCTGAGCCCTGCAGTGCAGAGAGAAGACAGCTATGCTTAAAGAAAGA				human.			
OY	1381	gagtgtaggggtgcccttcagtggtgctggtgccggtctaccagaagaatggtcat				human.			
Db	1279	TAAATGAGCGATGAGAGTG-----TGGGTTCATATGACGCTACCGGAGAGGCGAAC				human.			
OY	1441	gggaactgcccaaaatgtgcatgttgctgactacaagtgtgactgactgtaaggaacaattca				human.			
Db	1328	GGGAATATGTGAAGATGT-CTTTTGGCTTACAGCGGATGGATTGCCAAGACCAATTTC				human.			
OY	1501	gctgactcctcaattattgtgagcaccatcgctgtgcatgtgcatgtcattcgaatgataat				human.			
Db	1387	GCTGATCCTCAACA--TTGTGGAGAACCATGTGCTGAGCTTTCATCCTCATCTTGTCTAAT				human.			
OY	1561	gcatgattgttctactagcaagaatcaataaacaagaagcatatgtgaagaagagac				human.			
Db	1445	GTTTTATTCGTCTCATAGTCAAGTCAAAAGAACAAAGAAAGAGAGGA--GAAGAGCAGAT				human.			
OY	1621	ttgattgcagaaagacttcaaatctaactgagtcgacagagcttcaccacatctgt				human.			
Db	1502	CTGATTGGAGATGACTTTTCATTAACCTGGCAATGAGG-CCAACCGGCTTCTCCAA-CTTCG				human.			
OY	1661	gagcaataacgagcgctctcctcagtcgaagatcacgcctcca				human.			
Db	1560	GAGCAGACACGACGATCTTCCCAAAAGTCAAAACAGGGGTCCCGA				human.			
RESULT	12								
LOCUS	HSMB01289	1326 bp	mRNA	PRI	18-FEB-2000				
DEFINITION	Homo sapiens mRNA; CDNA DKFZ586J1521 (from clone DKFZ586J1521).								
ACCESSION	AL122124								
VERSION	AL122124.1	GI:6102978							
KEYWORDS									
ORGANISM	human.								
SOURCE	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
AUTHORS	1 (bases 1 to 1326)								
	Ottensmieder, B., Obermaier, B., Mewes, H.W., Gaassenhuber, J. and Wiemann, S.								
TITLE	Direct Submission								
JOURNAL	Submitted (15-OCT-1999) MIPs, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY								
COMMENT	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ). Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the CDNA sequencing consortium of the German Genome Project. This clone (DKFZ586J1521) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <a href="http://www.mips.biochem.mpg.de/proj/cdna/">http://www.mips.biochem.mpg.de/proj/cdna/</a> .								
FEATURES									
source	1..1326								
	/organism="Homo sapiens"								
	/db_xref="taxon:9606"								
	/clone="DKFZ586J1521"								
	/clone_lib="586 (synonym: hntel). Vector psportl; host DH10B; sites NotI + SalI/MluI"								
	/dev_stage="adult"								
	/tissue_type="uterus"								
	1299..1304								

PolyA-site	1307
BASE COUNT	432 a 190 c 220 g 484 t
ORIGIN	
Query Match	4.7%; Score 149.6; DB:9; Length 1326;
Best Local Similarity	97.4%; Pred. No. 3.1e-27;
Matches 152; Conservative	0; Mismatches 4; Indels 0; Gaps 0;
OY	5 gctcagatgtaaacatgcgaagaagtcaattacttcgttagagatttgcgtgacctaaga 64       Db 790 GATCCTGTGTTAAACTGCCACGAAGAAGTAATTACTCTGACGAGTTCGTGACCTTGAA 849     OY 65 gtgaacctgtttgtaaatgatgccctgatacataaaaccgagccatcatatctcca 124     Db 850 GTGAACAAGCTGTTGTAAGTAGAGCCCTGATCATATAAACGGACACAGGCATATTATCTCA 909     OY 125 agtgttaatatactgaacttatgcaatattcaaca 160       Db 910 AGTGTTAATATACCTGACTTATGCACTATTCAAACA 945 
RESULT 13	
AB023198	
LOCUS	AB023198 5182 bp mRNA PRI 16-JUN-1999
DEFINITION	Homo sapiens mRNA for KIAA0981 protein, partial cds.
ACCESSION	AB023198
VERSION	AB023198.1 GI:4589605
KEYWORDS	
SOURCE	'Homo sapiens adult male brain cdNA to mRNA, clone_1lb:pluescriptII SK plus clone:hj07094.'
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (sites)
TITLE	Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Hirotsawa,M., Miyajima,N., Tanaka,A., Kotani,H., Nomura,N. and Ohara,O., Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
JOURNAL	DNA Res. 6 (1), 63-70 (1999)
MEDLINE	99246063
REFERENCE	2 (bases 1 to 5182)
AUTHORS	Ohara,O., Nagase,T. and Kikuno,R.
TITLE	Direct Submission
JOURNAL	Submitted (04-FEB-1999) to the DDBJ/EMBL/Genbank databases. Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)
FEATURES	location/Qualifiers
Source	1..5182
gene	/organism="Homo sapiens"
CDS	/db_xref="taxon:9606" /clone_hjb="hj07094" /clone_lib="pluescriptII SK plus" /dev_stage="adult" /sex="male" /tissue_type="brain" 1..1738 /gene="KIAA0981"  1..1738 /gene="KIAA0981" /codon_start=2 /product="KIAA0981 protein" /protein_id="BAAF6825.1" /db_xref="GI:4589606" /translation="PSVSPFGRLRQGEESKISAMDASPRNISPGLONGEKEDREFLTTLSOSSTSHLOLPPEPVMSQSVGAPELDTASSSEDEVDFGLGSTDOVKERS TKMKIPFNILPDGNSYNPILPPFDKRIWAEHRVRIVACEKPSIITAPLSCKRY RNALDELSKATONHSAEEGLPTNTSSSRKRSSP ILHPKSGSQTRRTETEPQPK KASOMSLPFRETAKSDPLSSQRKETLRGADSVAYOVGCGKESTENQGVPEDEVG GDTDKOLINPHVELQGSDAWAKVCYLTYAGEFHKKREVILLDSSEDDFIRSLSSHSP

MWARGKSGAATPATEDDREIFLKMPRELVOSFLDPAPIHFNITNAVOCKRTALAK  
 ILGVNIRIGKNSQNTNEKIDLLVMENLEFGRMQAVFDLKSILRRNRKTDGKESC  
 DVLGIDENLKKNSQNTNEKIDLLVMENLEFGRMQAVFDLKSILRRNRKTDGKESC  
 ELVVGIIIDIRFTTMDKLEMYVKSTGILGGQCKMPTVSPELYRRFRFCBAMKYFLM  
 VPDHMTGLGINC"

BASE COUNT 1606 a 888 c 1045 g 1643 t  
 ORIGIN

Query Match 4.7%; Score 149.6; DB 9; Length 5182;  
 Best Local Similarity 97.4%; Pred. No. 3.4e-27;  
 Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 gctcgagttaaaactgccaagaaagtaattacctgtagagttgctgagcttgaaga 64  
 4653 GATCCTGTGTAAACCTGCCAAGAAAGTAATTACCTGTAGAGTTGCTGAGCTTGAAGA 4712

OY 65 gtgaaaaactgttgtaatgagctgatacaataaagcagccatcattatctctca 124  
 4713 GTGAAAACCTGTGTGAATGAGCCGTGATCATPAAACGGACGCCATTATTCCTCA 4772

OY 125 agtctaatatactgactatgacgatltcaacaaca 160  
 4773 AGTCTTAATATACGACTTATGCAAGTATTCACCAACA 4808

RESULT 14  
 AC012362 232816 bp DNA HTG 30-JUL-2001  
 LOCUS Homo sapiens chromosome 2 clone RP11-404D15, WORKING DRAFT  
 DEFINITION  
 AC012362  
 AC012362  
 AC012362.6 GI:15029477  
 HTG: HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 232816)  
 Waterston, R.H.  
 The sequence of Homo sapiens clone  
 unpublished  
 2 (bases 1 to 232816)  
 Waterston, R.H.  
 Direct Submission  
 Submitted (25-OCT-1999) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA  
 On Jul 30, 2001 this sequence version replaced gi:14572170.

COMMENT  
 ----- Genome Center -----  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: http://genome.wustl.edu/gsc/index.shtml  
 ----- Project Information -----  
 Center project name: H\_NH0404D15  
 ----- Summary Statistics -----  
 Sequencing vector: p13; 368  
 Sequencing vector: plasmid; 648  
 Chemistry: Dye-primer ET; 368 of reads  
 Chemistry: Dye-terminator Big Dye; 648 of reads  
 Assembly program: Phrap; version 0.990319  
 Consensus quality: 232390 bases at least Q40  
 Consensus quality: 232626 bases at least Q30  
 Consensus quality: 232815 bases at least Q20  
 Insert size: 226000; agarose-1p  
 Insert size: 232816; sum-of-ctnigs  
 Quality coverage: 8.23 in Q20 bases; agarose-1p  
 Quality coverage: 8.70 in Q20 bases; sum-of-ctnigs  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 ctnigs. The true order of the pieces

\* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the ctnigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 232816: ctnig of 232816 bp in length.  
 Location/Qualifiers  
 1. 232816  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /clone="RP11-404D15"  
 /note="assembly\_name:Contigs  
 clone\_end:SP6  
 vector\_side:right"

misc\_feature  
 1. 232816  
 /note="assembly\_name:Contigs  
 clone\_end:SP6  
 vector\_side:right"

BASE COUNT 67913 a 42697 c 46345 g 75861 t  
 ORIGIN

Query Match 4.7%; Score 149.6; DB 2; Length 232816;  
 Best Local Similarity 97.4%; Pred. No. 4.2e-27;  
 Matches 152; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 5 gctcgagttaaaactgccaagaaagtaattacctgtagagttgctgagcttgaaga 64  
 110309 GATCCTGTGTAAACCTGCCAAGAAAGTAATTACCTGTAGAGTTGCTGAGCTTGAAGA 110368

OY 65 gtgaaaaactgttgtaatgagctgatacaataaagcagccatcattatctctca 124  
 110369 GTGAAAACCTGTGTGAATGAGCCGTGATCATPAAACGGACGCCATTATTCCTCA 110428

OY 125 agtctaatatactgactatgacgatltcaacaaca 160  
 110429 AGTCTTAATATACGACTTATGCAAGTATTCACCAACA 110464

RESULT 15  
 MUSANT10A 2968 bp mRNA ROD 12-JUN-1993  
 LOCUS Mouse cell surface antigen 114/A10 mRNA, complete cds.  
 DEFINITION  
 J04634.1 GI:191943  
 VERSION  
 J04634.1 GI:191943  
 KEYWORDS  
 cell surface antigen.  
 Mouse (strain C57Bl6) adult bone marrow hemopoietic cell line  
 B6StuA, cDNA to mRNA, clones CDW8.A10.[1,2].  
 SOURCE  
 Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 2968)  
 Dougherty, G.J., Kay, R.J. and Humphries, R.K.  
 Molecular cloning of 114/A10, a cell surface antigen containing  
 highly conserved repeated elements, which is expressed by murine  
 hemopoietic progenitor cells and interleukin-3-dependent cell lines  
 J. Biol. Chem. 264, 6509-6514 (1989)

JOURNAL  
 MEDLINE  
 COMMENT  
 Draft entry and computer-readable sequence for [1] kindly submitted  
 by R.K. Humphries, 24-MAR-1989.

FEATURES  
 source  
 1. 2968  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 <1. 2111  
 /note="cell surface antigen mRNA (alt.)"  
 <1. 2162  
 /note="cell surface antigen mRNA (alt.)"  
 <1. 2268  
 /note="cell surface antigen mRNA (alt.)"  
 <1. 245  
 /note="cell surface antigen 114/A10 signal peptide"  
 195. 1916  
 /note="cell surface antigen 114/A10 precursor"



```
/codon_start=1
/protein_id="AAA37239.1"
/db_xref="GI:309106"
/translation="MKGFLLISLLLVTVSSSSQASSTSSSGCTSPPTVQSQSPG
SSQASTTSSSGASPPPTVQSQSPGSSQASTTSSSGASPPPTVQSQSPGSSQ
ASTTSSSGASPPPTVQSQSPGSSQASTTSSSGASPPPTVQSQSPGSSQASTT
TSSSGASPPPTVQSQSPGSSQASTTSSSGASPPPTVQSQSPGSSQASTT
GASSIVPSSGSGTSPDLNPNPCIKSKHPCLEGEYXNSLSQVKG
TFPDIMSVSETANLEDENSVGYOELNYSVDFEFTFNKTDGOTYLIKVSAPSR
SARSAMDATKDYVSIVNFGADTKETKESVSAIETATISNVDYISINCDH
GCVGNDSSKODILIOCTCKPOLDLNPOVPCVAVTSCQPCNAEKEQCLMDGVM
CVCMPGYORANGNRCECPFGYSQMNCKDQPOLILIVGIAIALILILIAIVSA
RSKNKKDGEORLIEDDFHNLRLRQTFNSLMGADNSIFPKVFTGVSQTPNPYANOR
SMRPDY"
repeat_region
243..883
/note="Ser-Thr tandem repeat region"
mat_peptide
246..1913
/note="cell surface antigen 114/A10"
polya_signal
2091..2097
polya_signal
2951..2956
BASE COUNT
783 a 756 c 707 g 722 t
ORIGIN
Query Match 4.5%; Score 144; DB 10; Length 2968;
Best Local Similarity 59.2%; Pred. No. 8.8e-26;
Matches 568; Conservative 0; Mismatches 345; Indels 47; Gaps 17;
QY 789 ataatacaaglttttgcctgtgtttagaagggtactactacaactctcta-----cat 842
Db 949 ATAGCAAAACACTCTGCTGTGTAGAAAGGCTATTACTACAACTGCTCTGTGCTCAT 1008
QY 843 gtaagaagaagaagttccctcgtgagaagatttcagtgacagttacagaacatttgac 902
Db 1009 GTGTGAAGGAAACGACATTTCCCTGGGACATTAGCA-TGAGTGTCTGTAACAGCCAAC 1067
QY 903 ccagaagaagaacattccatgacctcaagaactgtcatagtgaaattactagctgttt 962
Db 1068 TTGGAAGATGAACACTCTGTGGGCTATCAGAGTGTACAACTGTTACCGATTTT 1127
QY 963 aaagatgtatttggcacactcgtttagaacagactgtlaattctactcgt-----aagc 1017
Db 1128 GAAACACATTTAATAAGACTGACTATGACAGACTGTAATTATTAAGTGACACAGCC 1187
QY 1018 acatctcgtcaccaagaattcgaatcgtgctgtatgacaagtttgttaattgaca 1077
Db 1188 CTTTCACGCTCAGCAGAG-TCTGCATGCGTGAT--GCTACAAAGCAGCTCAGTATCC 1244
QY 1078 atagtaacaatttggcagaacaacacaagtgacaatgagaagactgtgactgagaat 1137
Db 1245 GTAGTAATATATTGTGGACGACACAAAGAACTGAAGAGTATCTTCTGCATTC 1304
QY 1245 gtatgaataatatttggcagaacacaaagaaactgaaagactgtatcttgcataac 1304
QY 1138 taataaagaacattataagtagctcaagaacatttctaactatgattggaacctgtcg 1197
Db 1305 GAA-----ACTGCATTTAAGACTGATGCAATGTGAAGACTATGTCAGTATAATTG 1358
QY 1198 gtgtgtgtgtgtttagaggtcgtggaaccaagactgtgtgtagactgtcctaagtgtt 1257
Db 1359 TGTGATCAATTATGTTGTGTAGGGGAAATGACAGCAATAATGCCAGACATTTTACAAATGC 1418
QY 1258 gcatgcatgtgtcaaatgtcactcgtcaaaagcccaaccacagagcccttctcgttg 1317
Db 1419 ACGTGC-----AAACCTGGGCTGGACAGGCTGAACCCACAGGTTCTTTTGTGTTG 1470
QY 1318 ctccagctcaagagtgctcgtatgctgcaacgacacagaagaagcgaatgcttaataa 1377
Db 1471 CT-----GTAACGTGCTCTCAGCCCTGCAATGCAGAGAGAGAG-GAGCAGTCTTAA 1522
QY 1378 gaagaagtgtgtgtccctcgaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1437
Db 1523 GATGATTAACGAGATTAATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1580
QY 1438 aatgggaactgccaaagtgtgactttaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1497
```

```
11 1111 11 1111 1111 1111 1111 11 11 11 11 1111
Db 1581 AACAGAAAGTGAAGAGATGTCATTT-GGCTACAGCGGAGATGATTCGAAGACCAATT 1639
QY 1498 tcaactgatccactacttatttgtggcaaccatcgctgcatgttcatctcaagcatgata 1557
Db 1640 TCAGCTGATCTCCAC--CATCGTGGGACCAATTGCTGGAGCCCTCATCTTGTGCTG 1697
QY 1558 attgattgatgttactagaagaatcaaatgacaagaagcgaagatatltgaagaagag 1617
Db 1698 ATTGCTTTATCTCTCT-GCGAGGTCAAGAACAAA-AGAAGGATGCGAAGAGC-AG 1754
QY 1618 aactgtatgacgaagacttccaataatcctaaactgcgtgcacaggtctcaccatct 1677
Db 1755 AGGCTGATTGAAGATGACTTCCATTAACCTAGCAGCTGAGG-CAGACCGCTTCTCCAA-CT 1812
QY 1678 atgagcatatcagcagcgtcttcctcctcagtgtaagattacggctccaagaccgcta 1737
Db 1813 TAGGAGCGGACAAACAGCATCTTCCCAAAAGTCAGAAAGGGGTCCCGAGTCAAGACCCCTA 1872
```

Search completed: March 30, 2002, 13:34:39  
Job time: 23561 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:31:07 ; Search time 5438.31 Seconds  
(without alignments)  
636.253 Million cell updates/sec

Title: US-09-867-034-2  
Perfect score: 322  
Sequence: 1 gacacgaacacacccctga.....ctccaagtgtcgaagact 322

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hlc:\*  
10: gb\_estl:\*  
11: gb\_est2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_liv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	321	99.7	441	10	AM364247 OV3-DT001
C 2	321	99.7	484	11	BF814154 RC3-CI004
C 3	321	99.7	489	11	BF513917 UR-H-BM1
C 4	321	99.7	619	10	AM161619 x146D06.x
C 5	321	99.7	623	11	BE867131 BE867131
C 6	321	99.7	637	10	AM364300 OV3-DT001
C 7	321	99.7	666	11	BG778248 BG778248
C 8	321	99.7	689	10	AM582256 OV4-ST021
C 9	321	99.7	689	10	BE748141 BE748141
C 10	321	99.7	721	10	AM956284 EST368354
C 11	321	99.7	734	11	BE870718 BE870718
C 12	321	99.7	742	11	BG823844 BG823844

13	321	99.7	750	11	BG179160	BG179160 602330032
14	321	99.7	751	11	BG386151	BG386151 602455333
15	321	99.7	853	11	BG610804	BG610804 602612442
16	321	99.7	890	11	BG259927	BG259927 602371889
17	321	99.7	948	11	BE796856	BE796856 601588166
C 18	320	99.4	674	10	AI075324	AI075324 ov20a02.x
C 19	320	99.4	708	10	AI675865	AI675865 wB97h01.x
C 20	320	99.4	866	11	BG285246	BG285246 602409569
21	319.4	99.2	755	11	BF733142	BF733142 EST039 HU
C 22	319	99.1	595	10	AA573742	AA573742 HK07a09.S
C 23	319	99.1	744	10	AI815198	AI815198 WK72C03.x
C 24	319	99.1	782	10	AL543432	AL543432 AL543432
C 25	317.8	98.7	450	10	BE072059	BE072059 PM4-BT053
C 26	317.8	98.7	503	11	BF155131	BF155131 PM3-BT083
C 27	317.8	98.7	627	10	AM369331	AM369331 OVI-BN000
28	317	98.4	537	10	AA314225	AA314225 EST186091
C 29	316.8	98.4	643	10	AM151674	AM151674 x167e12.x
C 30	316.4	98.3	547	11	BE841679	BE841679 MR1-SN006
C 31	314.6	97.7	459	11	BF798311	BF798311 RC3-CI004
C 32	314.6	97.7	672	10	AI634717	AI634717 t209d10.x
C 33	312.2	97.0	566	10	AA316115	AA316115 EST187838
C 34	310.6	96.5	608	10	AI436796	AI436796 t182905.x
C 35	309.8	96.2	700	10	AI817063	AI817063 w176e07.x
C 36	309	96.0	740	11	BG501149	BG501149 602546389
C 37	308.8	95.9	659	10	AI380637	AI380637 t402d07.x
C 38	307.4	95.5	669	10	AI888294	AI888294 WK30e06.x
C 39	306.4	95.2	609	10	AM614754	AM614754 h138a09.x
C 40	306	95.0	592	10	AA909945	AA909945 OM32908.S
C 41	305.4	94.8	551	10	AA307697	AA307697 EST178713
C 42	305	94.7	403	10	AM192785	AM192785 x151c07.x
C 43	303	94.1	631	10	AI080640	AI080640 OX54c07.x
C 44	303	94.1	682	10	AM194118	AM194118 km12d06.x
C 45	302.6	94.0	475	11	BF744154	BF744154 PM3-BT083

#### ALIGNMENTS

RESULT 1  
LOCUS AM364247 441 bp mRNA 04-FEB-2000  
DEFINITION OV3-DT0012-081299-021-e06 DT0012 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM364247  
VERSION AM364247.1 GI:6868897  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS HCSP <http://www.ludwig.org.br/ORESTES>.  
TITLE The FAPESP/LICR Human Cancer Genome Project  
JOURNAL Unpublished (1999)  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=OV3&st=QV3-DT0012-081299-021-e06&t3=1999-12-08&t4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 7  
High quality sequence stop: 440.

#### FEATURES

Source location/Qualifiers  
1..441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="DT0012"

```
/dev_stage="Adult"
/notes="Organ: denis_drash; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

BASE COUNT 114 a 90 c 96 g 141 t

ORIGIN

Query Match 99.7%; Score 321; DB 10; Length 441;  
Best Local Similarity 99.7%; Pred. No. 8.7e-78;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gaaagcaaaaccccttgatgattatcacccttgatgagtgccacacagtcagc 60  
|||||  
DB 424 GACAGACAAACCCCTTGATGATTTATTCACCTTGATGAGTGCACACAGTCAAGC 365  
|||||

OY 61 ttcaagaaagtgcttgctgaataaagaatccagaatctgcaagacgttgctc 120  
|||||  
DB 364 TTTAAAGAAAGTGTGCTGATAAATAAGAAATCCAGAAATTCGACAGAGTTTGCT 305  
|||||

OY 121 cctcaatctggttatagaacaactgcaaacacctcttcctgatyggcagatgtccc 180  
|||||  
DB 304 CCTCAATCTGTTTATGAAACAACCTGACAAACACCTTCTCTGATGGCCAGTATGTC 245  
|||||

OY 181 cagattatgttggtgacccatctcgcagattagagccgatatcactggaagatctc 240  
|||||  
DB 244 CAGGATTATGTTGTTGACCCATCTCGACAGTTAGCCGATATACCTGGAAGATATTC 185  
|||||

OY 241 aaancgtctctatgcttaagaacctgcagatacagctctgttgcttacaacatgaaga 300  
|||||  
DB 184 AATTCGCTCTATGCTTACGACAACTCGAGATACAGCTCTGTGCTTGACAACTGAAGA 125  
|||||

OY 301 agctctcaagtgctgaagact 322  
|||||

DB 124 AGCTCTCAAGTGTCTGAAGACT 103  
|||||

RESULT 2

BF814154 484 bp mRNA EST 12-JAN-2001

LOCUS BF814154

DEFINITION RC3-C10043-281100-025-b10 C10043 Homo sapiens cDNA, mRNA sequence.

ACCESSION BF814154

VERSION BF814154.1 GI:12146229

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

1 (bases 1 to 484)

Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R.,  
Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F.,  
Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H.,  
Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare,  
M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and  
Simpson, A. J.

Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the PAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL

```
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC3&l2=RC3-C10043-  
.281100-025-b10&l3=2000-11-28&l4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14  
High quality sequence stop: 484.  
Location/Qualifiers
```

FEATURES

source

1. 484

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="C10043"

/dev\_stage="Adult"

/notes="Organ: colon\_ins; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the pUC 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT 148 a 114 c 95 g 127 t

ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 484;  
Best Local Similarity 99.7%; Pred. No. 8.7e-78;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 gacaagcaaaaccccttgatgattatcacccttgatgagtgccacacagtcagc 60  
|||||  
DB 97 GACAGACAAACCCCTTGATGATTTATTCACCTTGATGAGTGCACACAGTCAAGC 156  
|||||

OY 61 ttcaagaaagtgcttgctgaataaagaatccagaatctgcaagacgttgctc 120  
|||||  
DB 157 TTTAAAGAAAGTGTGCTGATAAATAAGAAATCCAGAAATTCGACAGAGTTTGCT 216  
|||||

OY 121 cctcaatctggttatagaacaactgcaaacacctcttcctgatyggcagatgtccc 180  
|||||  
DB 217 CCTCAATCTGTTTATGAAACAACCTGACAAACACCTTCTCTGATGGCCAGTATGTC 276  
|||||

OY 181 cagattatgttggtgacccatctcgcagattagagccgatatcactggaagatctc 240  
|||||  
DB 277 CAGGATTATGTTGTTGACCCATCTCGACAGTTAGCCGATATACCTGGAAGATATTC 336  
|||||

OY 241 aaancgtctctatgcttaagaacctgcagatacagctctgttgcttacaacatgaaga 300  
|||||  
DB 337 AATTCGCTCTATGCTTACGACAACTCGAGATACAGCTCTGTGCTTGACAACTGAAGA 396  
|||||

OY 301 agctctcaagtgctgaagact 322  
|||||

DB 397 AGCTCTCAAGTGTCTGAAGACT 418  
|||||

RESULT 3

BF513917/c 489 bp mRNA EST 07-DEC-2000

LOCUS BF513917/c

DEFINITION UI-H-BW1-amq-f-02-0-UI.s1 NCI-CGAP\_Sub7 Homo sapiens cDNA clone

ACCESSION BF513917

VERSION BF513917.1 GI:11599096

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

1 (bases 1 to 489)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov

COMMENT The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A



```

Db 317 CAGGATTATTTGTTGATGACCATCTCTGACAGTTAGAGCCGATATCATCTGGAAGATATTC 238
QY 241 aaangctctatagcttacgaacctgcagatacagctctgtgtcttacacaatgaagaa 300
Db 257 AATTCGTCTCTATGCTTACGACACCTGCAGATACAGCTCTGTGCTTGACAACTGAAGAA 198
QY 301 agcttcaagttgctgaagact 322
Db 197 AGCTCTCAAGTCTGTAAGACT 176

RESULT 5
BE667131 623 bp mRNA EST 20-OCT-2000
LOCUS 601442973F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3846834 5'
DEFINITION mRNA sequence.
ACCESSION BE667131
VERSION BE667131.1 GI:10315907
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE NIH-MGC http://mgi.mci.nih.gov/
AUTHORS 1 (bases 1 to 623)
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM9560 row: b column: 19
High quality sequence stop: 621.
Location/Qualifiers
1..623
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 198 a 149 c 126 g 150 t
ORIGIN
Query Match 99.7%; Score 321; DB 11; Length 623;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaacaaccccttgatgatattatcatcactgtagtgagccacacagtaaac 60
Db 194 GACACGACCAACACCTTATGATTTATTCATCTCTGATGAGTGCACACAGTCAAGC 253
QY 61 tttaagaagaagtgtgtcgtgaataataagaatccagaatctgcagagacgtttgctt 120
Db 254 TTTAAAGAAAGTGTGCGAAGAAATTAAGAAATCCAGAAATTCGACAGAGAGTTGTGCT 313
QY 121 cctcaatctggtttatgaacaactgcaacaacaccttctcctgtaggccaagtatgcc 180
Db 314 CCTCAATCTGTTTATGAACAACATGACAAACACCTTCTCTGATGAGCCAGTATGTCCC 373
QY 181 caggaatgatgtttagccacctctgcacgttagagccgatatatacaggaagatattc 240

```

```

Db 374 CAGGATTATTTGTTGATGACCATCTCTGACAGTTAGAGCCGATATCATCTGGAAGATATTC 433
QY 241 aaangctctatagcttacgaacctgcagatacagctctgtgtcttacacaatgaagaa 300
Db 434 AATTCGTCTCTATGCTTACGACACCTGCAGATACAGCTCTGTGCTTGACAACTGAAGAA 493
QY 301 agcttcaagttgctgaagact 322
Db 494 AGCTCTCAAGTCTGTAAGACT 515

RESULT 6
AM364300/c 637 bp mRNA EST 04-FEB-2000
LOCUS QV3-DT0012-291299-051-g07 DT0012 Homo sapiens cDNA, mRNA sequence.
DEFINITION QV3-DT0012-291299-051-g07 DT0012 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM364300
VERSION AM364300.1 GI:6868950
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE HCGP http://www.ludwig.org.br/ORESTES
AUTHORS The FAPESP/LICR Human Cancer Genome Project
TITLE Unpublished (1999)
JOURNAL Contact: Simpson A.J.G.
COMMENT Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l-QV3&c2-QV3-DT0012-
291299-051-g07&t3-1999-12-29&t4-1)
Seq primer: puc 18 forward
High quality sequence stop: 50.
High quality sequence stop: 609.
Location/Qualifiers
1..637
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="DT0012"
/dev_stage="Adult"
/notes="Organ: denis_drash; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 151 a 131 c 155 g 200 t
ORIGIN
Query Match 99.7%; Score 321; DB 10; Length 637;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaacaaccccttgatgatattatcatcactgtagtgagccacacagtaaac 60
Db 467 GACACGACCAACACCTTATGATTTATTCATCTCTGATGAGTGCACACAGTCAAGC 408
QY 61 tttaagaagaagtgtgtcgtgaataataagaatccagaatctgcagagacgtttgctt 120
Db 407 TTTAAAGAAAGTGTGCGAAGAAATTAAGAAATCCAGAAATTCGACAGAGTGTGCTCT 348
QY 121 cctcaatctggtttatgaacaactgcaacaacaccttctcctgtaggccaagtatgcc 180
Db 347 CCTCAATCTGTTTATGAACAACATGACAAACACCTTCTCTGATGAGCCAGTATGTCCC 288

```

QY 181 cagattatgtttgttgcacccatctctacagattagagccatatactactggaagaatc 240  
|||||  
Db 287 CAGATTATGTTGTGTGACCCATCTCTACAGATTAGAGCCCATATCACTGGAAAGTATTTC 228  
QY 241 aaangctctatgtcttgcagaaactgcagatacagctctgttgccttgacaacatgaaga 300  
|||||  
Db 227 AAATGCTCTATGCTTATGCAAGCTGCAGATACAGACTCTGTGCTTGACACATGAAGAA 168  
QY 301 agctctcaagttgctgaagact 322  
|||||  
Db 167 AGCTCTCAAGTTGCTGAAGACT 146

RESULT 7  
LOCUS BG778248 666 bp mRNA EST 15-MAY-2001  
DEFINITION 602666830F1 NIH\_MGC\_60 Homo sapiens cDNA clone IMAGE:4806679 5',  
mRNA sequence.  
ACCESSION BG778248  
VERSION BG778248.1 GI:14048565  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE NIH-MGC http://mgc.nci.nih.gov/  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgaabs-remail.nih.gov  
Tissue Procurement: DCTD/DRP

CDNA Library Preparation: CLONTECH Laboratories, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.lnl.gov  
plate: LCM1656 row: 1 column: 08  
High quality sequence stop: 665.  
Location/Qualifiers

## FEATURES

source

1. 666  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4806679"  
/clone\_id="NIH\_MGC\_60"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (T1 phage-resistant)"  
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);  
Site\_1: SfiI (ggcgctcgcc); Site\_2: SfiI (ggcgctatggcc  
5' and 3' adaptors were used in cloning as follows: 5'  
adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor  
sequence: 5'-ATCTAGAGCGCGCGCCGACATG-dT(30)BN-3'  
(where B = A, C, or G and N = A, C, G, or T). Average  
insert size 1.5 kb (range 0.9-4.0 kb). 14/15 colonies  
contained inserts by PCR. This library was enriched for  
full-length clones and was constructed by Clontech  
Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
Library."

BASE COUNT

205 a 165 c 140 g 156 t

ORIGIN

Query Match

99.7%; Score 321; DB 11; Length 666;

Best Local Similarity 99.7%; Pred. No. 8.7e-78;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagacaacaccccttgatgatattatcatcacttgatgagccacacagtcacgc 60  
|||||  
Db 269 GACAAGACAACCCCTTGATGATTATTCATCACTTGATGATGTCACACACAGTCAAGC 328

QY 61 tttaagaaagtgttgcgtgaataaagaaatccagaattggcagagcagttgtcct 120  
|||||  
Db 329 TTTAAAGAAAGTGTGTGCTGAAAAATTAAGAAATCCAGAAATTTGGCAGACGATTGTCTT 388  
QY 121 cctcaatcgtgttataagaaactgacaacacaccttctcctgaltggtccaglatgtccc 180  
|||||  
Db 389 CCTCAATCTGTTATGTAACAACTGACAAACACCTTCTCCTGATGGCCAGTATGTCCTC 448  
QY 181 cagattatgttgcgtgacccatctctgcagattagagccgatalactatgcgaatattc 240  
|||||  
Db 449 CAGATTATGTTGTGTGACCATCTCTGACAGTTTAGACCCATATCACTGGAAAGTATTTC 508  
QY 241 aaangctctatgtcttgcagaaactgcagatacagctctgttgccttgacaacatgaaga 300  
|||||  
Db 509 AAATGCTCTATGCTTATGCAAGCTGCAGATACAGCTCTGTGCTTGACACATGAAGAA 568  
QY 301 agctctcaagttgctgaagact 322  
|||||  
Db 569 AGCTCTCAAGTTGCTGAAGACT 590

RESULT 8  
LOCUS AM582256 689 bp mRNA EST 16-MAR-2000  
DEFINITION OVA-ST0212-120100-075-e10 ST0212 Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM582256  
VERSION AM582256.1 GI:7257305  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE HCGP http://www.ludwig.org.br/ORESTES.  
AUTHORS The FAPESP/LICR Human Cancer Genome Project  
TITLE Unpublished (1999)  
JOURNAL Contact: Simpson A.J.G.  
COMMENT Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV4&tl2=QV4-ST0212-  
120100-075-e10&tl3=2000-01-12&tl4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 687.  
Location/Qualifiers

## FEATURES

source

1. 689  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="ST0212"  
/dev\_stage="Adult"  
/note="Organ: stomach; Vector: puc18; Site\_1: SmaI;  
Site\_2: SmaI; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

BASE COUNT

212 a 172 c 143 g 162 t

ORIGIN

Query Match

99.7%; Score 321; DB 10; Length 689;

Best Local Similarity 99.7%; Pred. No. 8.7e-78;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagacaacaccccttgatgatattatcatcacttgatgagccacacagtcacgc 60  
|||||

```

Db 282 GACAGCAACAAACCTTGATGATTATTCATCACTGGAGTGGTCCACACAGTCAAGC 341
QY 61 tttaagaagaagtgcttgatgaataaagaatccagaatctgagcagatgtgctt 120
342 TTTTAAAGAAAGTGTTCCTGTAATAAAGAAATCCAGAAATGGCAGAGCATTTGTCCT 401
QY 121 ccttaactggttatagaacaactgaacaacaccttctctcgtatgagcagatgtgcc 180
402 CCTCAATCTGTTTATGTAACCAACGACAAACACCTTTCCTCATGATGTCCTCC 461
QY 181 cagattatgttgtagccatctcgaagtgtagccgatatcctggaatatctc 240
462 CAGATTATGTTGTGTGACCATCTGACAGTTAGAGCCGATATCACTGGAATATTC 521
QY 241 aaangctctatgcttaagcaacctgagatagacgtctgtgcttgtaacaatagaaga 300
522 AAATGCTGTCTATGCTTACGACACTGCAGATACAGCTGTGCTTGACAAATGAAAGA 581
QY 301 agctctcaagttgctgaagact 322
Db 582 AGCTCTCAAGTTGCTGGAAGACT 603

RESULT 9
BE748141 689 bp mRNA EST 15-SEP-2000
LOCUS 601571509F1 NIH_MGC_55 Homo sapiens cDNA IMAGE:3838347 5',
DEFINITION mRNA sequence.
ACCESSION BE748141
VERSION BE748141
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC. http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM25 row: a column: 04
High quality sequence stop: 643.
Location/Qualifiers
1. 689
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_55"
/tissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/site_1="SfiI (ggcgctcgagc); Site_2="SfiI (ggcattatggc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATATATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGCGCGCATG-dT(30)BN-3',
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 209 a 170 c 148 g 162 t
ORIGIN

```

```

Query Match 99.7%; Score 321; DB 10; Length 689;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaaacccctgattatctatcacttgatgagtgccacacagtcagc 60
Db 251 GACAGCAACAAACCTTGATGATTATTCATCACTGGAGTGGTCCACACAGTCAAGC 310
QY 61 tttaagaagaagtgcttgatgaataaagaatccagaatctgagcagatgtgctt 120
Db 311 TTTTAAAGAAAGTGTTCCTGTAATAAAGAAATCCAGAAATGGCAGAGCATTTGTCCT 370
QY 121 ccttaactggttatagaacaactgaacaacaccttctctcgtatgagcagatgtgcc 180
Db 371 CCTCAATCTGTTTATGTAACCAACGACAAACACCTTTCCTCATGATGTCCTCC 430
QY 181 cagattatgttgtagccatctcgaagtgtagccgatatcctggaatatctc 240
Db 431 CAGATTATGTTGTGTGACCATCTGACAGTTAGAGCCGATATCACTGGAATATTC 490
QY 241 aaangctctatgcttaagcaacctgagatagacgtctgtgcttgtaacaatagaaga 300
Db 491 AAATGCTGTCTATGCTTACGACACTGCAGATACAGCTGTGCTTGACAAATGAAAGA 550
QY 301 agctctcaagttgctgaagact 322
Db 551 AGCTCTCAAGTTGCTGGAAGACT 572

RESULT 10
AW956284 721 bp mRNA EST 01-JUN-2000
LOCUS AW956284
DEFINITION EST368354 MAGE resequences, MAGD Homo sapiens cDNA, mRNA sequence.
ACCESSION AW956284
VERSION AW956284.1 GI:8145967
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 721)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
JOURNAL Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
COMMENT Unpublished (2000)
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: john@igfgr.org
Plate: 90
Seq primer: Reverse.
Location/Qualifiers
1. 721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGD"
/notes="Vector: pBluescriptSkm"
BASE COUNT 224 a 173 c 148 g 175 t 1 others
ORIGIN
Query Match 99.7%; Score 321; DB 10; Length 721;
Best Local Similarity 99.7%; Pred. No. 8.7e-78;
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 gacaagcaaaacccctgattatctatcacttgatgagtgccacacagtcagc 60
Db 260 GACAGCAACAAACCTTGATGATTATTCATCACTGGAGTGGTCCACACAGTCAAGC 319

```



QY	61	tttaagaagaagctcttgcctgaataataagaatccagaatctggcagagaagcttgcct	120
Db	320	TTTAAAGAAAGCTGTTTCTCTAAAAATTAAGAAATCCAGAAATTTGGCAGACAGTTTCTCT	379
QY	121	cctcaatctggtlthalyaanaacttgacaacaccttctctctgalygccagatgtccc	180
Db	380	CCTCAATCTGGTTTATGAACAACATGCACAACACTTTTCTCTGAGGCGCAGATATGCC	439
QY	181	caggattatgttgtaaccatctcgaagaattagccgatatatcagcggagaatattcc	240
Db	440	CAGATTATGTTTGTGTATCCCAATCTCGAAGATTAGCCGATATCATCTGGAAGATATTC	499
QY	241	aaanagctctatagtcttaacgaacctgagataaagctctgttgccttgaacaaactgaaga	300
Db	500	AAATGCTCTATGCTTACGAAACCTGCAGATACAGCTCTGTTGCTTGACAACTAATGAAGA	559
QY	301	agctctcaagttgctgaagact	322
Db	560	AGCTCTCAAGTGTGCTGAAGACT	581

RESULT	11
BE870718	
LOCUS	BE870718 734 bp mRNA
DEFINITION	60144857.0F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3852448 5'
ACCESSION	BE870718
VERSION	BE870718.1 GI:10319507
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 734)	NIH-MGC	<a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a>	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	Contact: Robert Strausberg, Ph.D.		

Email: [cgapbds-r@mail.nih.gov](mailto:cgapbds-r@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MCC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLM9574 row: 1 column: 17  
High quality sequence stop: 723.

FEATURES	Location/Qualifiers
source	1. .734

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3852448"
/clone_lib="NIH-MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: PCMV-SpOrtE; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dT."
Average insert size 1.8 kb. Library constructed by Life
Technologies. "

```

BASE COUNT	231 a	163 c	156 g	184 t
ORIGIN				

Query Match	99.7%	Score 321	DB 11	Length 734
Best Local Similarity	99.7%	Pred. No. 8.7e-78		
Matches 321	Conservative	0	Mismatches 1	Indels 0
				Gaps 0

0Y 1 gacagcaacaacacctgatgatattcactcttgatgagatgccacacgtcaagc 60  
|||||  
232 GACAGCAACAACACCTTGATGATTATTCACACTCTGGATGAGCGCCACACACTCAAGC 291  
Db  
0Y 61 tttaagaanaatgcttgcctgtaaaataagaataccagaataatgacgaagcagtttgcct 120

Accession	Sequence	Length
Db	TTTAAAGAAAGTGTGGTGAATAAAGAAATCCAGAAATTTGGCAGAGAGTGTGGCT	351
Qy	121 cctcaatcttgattatgtaaataacttgacaacaccttctccgtatgagccagtatgcc	180
Db	352 CCTCAATCTGGTTATGTAAGAACTGACAAACACCTTTCTCTGATGGCCAGATATGTC	411
Qy	181 caggaattatgtctgtgacccatctctgacagttagagccgatactcagtgaagatactc	240
Db	412 CAGGATTTATGTTGTTGGTACCCTCTGCAGCTTAGAGCCGATATCTACTGGAAATATTC	471
Qy	241 aaanagtcctatgcttaagaaactgcaagataaagcctctgtgttbaacaataagaa	300
Db	472 AAACCGTCTGATGCTTAAACGAACCTGGAGATACAGCTCTGTGGTTGACAAATGAA	531
Qy	301 agactctaaagtctgaagact 322	
Db	532 AGCTCTCAAGTTGCTGAAGACT 553	

[illegible]

REFERENCE	1 (bases 1 to 742)
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: ATCC  
CDNA library preparation: Ling Hong/Rubin laboratory  
CDNA library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCM1764 row: f column: 01  
High quality sequence stop: 702.

FEATURES  
source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4878336"
/clone_1ib="NH.MGC.15"
/tissue_type="epithelial carcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by Oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5
adapter: GGCACGAGG(G). Size-selected >500bp for average
insert size: 1.8kb. Library constructed by Ling Hong in

```

the laboratory of Gerald M. California, Berkeley) using

ORIGIN	Score	DB	Length
Query Match	99.7%	Score 321;	DB 11; Length 742;

```

Query Match      99.7%;  Score 321;  DB 11;  Length 742;
Best Local Similarity 99.7%;  Pred. No. 8.7e-78;
Matches 321;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps:
0y 1 gacagcacaacaccttgatgatattatcatcacttgatgagtcgcacacagctaacg 60
|||||

```

Db 239 GACACGACAAACCCCTTGATGATTATTCATCACTGGATGAGGCCACACAGTCAAGC 298  
Qy 61 tttaagaagaagtgctgtcgtgaataaagaataccagaatctgcagagcagttgtcct 120  
Db 299 TTTAAAGAAAGTGTTCCTGAAAATAAGAAATCCAGAAATGGCAGCAGTTGTGCT 358  
Qy 121 ccttaactcgtgtttaaagaacaactgacaacaccccttcctcgtgagtcagatgtccc 180  
Db 359 CCTCAATCTGGTTTATGAAACAACTGACAAACACCTTCTCCTGATGCGCCAGTATGTCC 418  
Qy 181 cagattatgtttgttgaaccatctcgaagcttagagccgatatcaactggaagatattc 240  
Db 419 CAGATTATGTTTGTTCACCATCTTGACAGTTAGAGCCGATATCACTGGAAGATATTC 478  
Qy 241 aaanagctctatgtcttaacgaacctgcagatacagctcgtgtgttcttgaacaatgaaga 300  
Db 479 AAATCGTCTGTATGCTTACGAACTGCAGATACAGCTGTGTGCTTGACAACTGAAGAA 538  
Qy 301 agctctcaagttgctgaagact 322  
Db 539 AGCTCTCAAGTTGCTGAAGACT 560

## RESULT 13

BG179160 750 bp mRNA EST 06-FEB-2001  
LOCUS 602330032F1 NIH\_MGC\_91 Homo sapiens cDNA clone IMAGE:4431583 5',  
DEFINITION mRNA sequence.  
BG179160  
ACCESSION BG179160.1 GI:12685943  
VERSION BG179160.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLES 1 (bases 1 to 750)  
JOURNAL NIH-MGC http://mgc.nci.nih.gov/  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10186 row: O column: 08  
High quality sequence stop: 732.

## FEATURES

## source

1. .750  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4431583"  
/clone\_lib="NIH\_MGC\_91"  
/tissue\_type="adenocarcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: prostate; Vector: pCMV-SPORT6; Site:1: NotI;  
Site:2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 1.4 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH-MGC Library."  
BASE COUNT 234 a 174 c 158 g 183 t 1 others  
ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 750;  
Best Local Similarity 99.7%; Pred. No. 8.7e-78;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gacacgaacaaaccccttgatgatatttcacactgagtgagtgccacacagtcgaagc 60  
|||||

Db 263 GACACGACAAACCCCTTGATGATTATTCATCACTGGATGAGGCCACACAGTCAAGC 322  
Qy 61 tttaagaagaagtgctgtcgtgaataaagaataccagaatctgcagagcagttgtcct 120  
Db 323 TTTAAAGAAAGTGTTCCTGAAAATAAGAAATCCAGAAATGGCAGCAGTTGTGCT 382  
Qy 121 ccttaactcgtgtttaaagaacaactgacaacaccccttcctcgtgagtcagatgtccc 180  
Db 383 CCTCAATCTGGTTTATGAAACAACTGACAAACACCTTCTCCTGATGCGCCAGTATGTCC 442  
Qy 181 cagattatgtttgttgaaccatctcgaagcttagagccgatatcaactggaagatattc 240  
Db 443 CAGATTATGTTTGTTCACCATCTTGACAGTTAGAGCCGATATCACTGGAAGATATTC 502  
Qy 241 aaanagctctatgtcttaacgaacctgcagatacagctcgtgtgttcttgaacaatgaaga 300  
Db 503 AAATCGTCTGTATGCTTACGAACTGCAGATACAGCTGTGTGCTTGACAACTGAAGAA 562  
Qy 301 agctctcaagttgctgaagact 322  
Db 563 AGCTCTCAAGTTGCTGAAGACT 584

## RESULT 14

BG386151 751 bp mRNA EST 12-MAR-2001  
LOCUS 60245333F1 NIH\_MGC\_15 Homo sapiens cDNA clone IMAGE:458336 5',  
DEFINITION mRNA sequence.  
BG386151  
ACCESSION BG386151.1 GI:13279597  
VERSION BG386151.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
TITLES 1 (bases 1 to 751)  
JOURNAL NIH-MGC http://mgc.nci.nih.gov/  
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCH1307 row: b column: 09  
High quality sequence stop: 672.

## FEATURES

## source

1. .751  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:458336"  
/clone\_lib="NIH\_MGC\_15"  
/tissue\_type="adenocarcinoma cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pORF7; Site:1: XhoI; Site:2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(C). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)"  
BASE COUNT 230 a 181 c 153 g 187 t  
ORIGIN

Query Match 99.7%; Score 321; DB 11; Length 751;  
Best Local Similarity 99.7%; Pred. No. 8.7e-78;  
Matches 321; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 gacacgaacaaaccccttgatgatatttcacactgagtgagtgccacacagtcgaagc 60  
|||||

Job time: 13928 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 07:01:58 ; Search time 8456.85 seconds

(without alignments)  
1777.131 Million cell updates/sec

Title: US-09-867-034-1

Sequence: 1 ttttttttttgcctgtttt.....cacgcgtggcctccggagcgc 911

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

GenEmbl: 1: gb\_da:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_com:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rnd:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rnd:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	607	66.6	57000	9 AB038162	AB038162 Homo sapi
2	607	66.6	157739	9 AP001623	AP001623 Homo sapi
3	607	66.6	181511	9 AC015555	AC015555 Homo sapi
4	607	66.6	340000	9 AP001746	AP001746 Homo sapi
5	516.8	56.7	834	6 AX017268	AX017268 Sequence
6	516.8	56.7	834	6 AX018086	AX018086 Sequence
7	403	44.2	480	6 AX193486	AX193486 Sequence
8	403	44.2	480	6 HUMPIBX	AX193486 Sequence
9	305.4	33.5	570	6 HUMPIBX	AX193486 Sequence
10	264.4	29.0	398	6 HUMPIBX	AX193486 Sequence
11	236.2	25.9	250	6 HUMPIBX	AX193486 Sequence
12	236.2	25.9	250	6 HUMPIBX	AX193486 Sequence
13	220.6	24.2	248	6 AX106296	AX106296 Sequence
14	220.6	24.2	248	6 AX140587	AX140587 Sequence
15	195	21.4	320	9 HSHIRF3	AX140587 Sequence
16	191	21.0	730	9 HSHIRF2	AX140587 Sequence
17	191	21.0	68531	2 AC055868	AC055868 Homo sapi
18	182	20.0	182	11 G06131	G06131 human STS W
19	182	20.0	182	11 G34583	G34583 human STS S
20	171	18.8	277	6 AX062739	AX062739 Sequence
21	134.4	14.8	206	6 AX062644	AX062644 Sequence
22	108.8	11.9	450	10 MUSMITF	D38410 Mouse mRNA
23	108.8	11.9	472	10 BC011042	BC011042 Mus muscu
24	108.8	11.9	7035	10 MMU271004	AJ271004 Mus muscu
25	105.2	11.5	424	10 RN048825	U48825 Rattus norv
26	103	11.3	184	10 AF012534	AF012534 Rattus no
27	101.2	11.1	5116	10 MMU46858	U46858 Mus musculu
28	100.6	11.0	264	10 RNIRF	X6956 R.norvegicu
29	100.6	11.0	264	10 S49317	S49317 Intestinal
30	100.4	11.0	431	10 RATTREOIL	M80826 Rat intesti
31	94.8	10.4	1128	9 HSHIRF1	U25654 Human intes
32	92.4	10.1	144	10 MTRFEX2	U24221 Mus musculu
33	64	7.0	6779	10 MMU271003	AJ271003 Mus muscu
34	63.6	7.0	574	4 SPSPPA	X51696 S.scrofa pa
35	61.8	6.8	546	5 XELXP2A	M90095 Xenopus lae
36	61.2	6.7	1471	5 XLPAPG	X51394 Xenopus mRN
37	60.2	6.6	447	10 MPPS2MPDP	Z21858 M.musculu
38	60	6.6	546	10 MMSPPA	X51697 M.musculu
39	60	6.6	8241	6 AX069291	AX069291 Sequence
40	59.6	6.5	92580	2 RN141N5	AL603725 Rattus no
41	59.2	6.5	5505	6 AX069290	AX069290 Sequence
42	58.8	6.5	466	10 RATS2P	D83231 Rattus norv
43	58.8	6.5	5142	6 AX069289	AX069289 Sequence
44	58.8	6.5	7048	10 MMU271002	AJ271002 Mus muscu
45	57.6	6.3	4405	10 MMU78770	U78770 Mus musculu

## ALIGNMENTS

RESULT 1  
AB038162/c  
LOCUS  
DEFINITION Homo sapiens TTF gene cluster for trefoil factor, complete cds.  
ACCESSION AB038162  
VERSION AB038162.1 GI:10280533  
KEYWORDS trefoil factor.  
SOURCE Homo sapiens pre-pro-B cell cell\_line:FLB14-14 DNA, clone\_lib:Keio  
BAC library clone:K8169B4.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (sites)

## REFERENCE

## AUTHORS

Berry, A., Scott, H.S., Kudoh, J., Tallor, I., Korostishevsky, M.,  
Wattenhofer, M., Guiponi, M., Barras, C., Rossier, C., Shinyu, K.,  
Wang, J., Kawasaki, K., Asakawa, S., Minoshima, S., Shimizu, N.,  
Antonarakis, S. and Bonne-Tamir, B.  
Refined localization of autosomal recessive nonsyndromic deafness  
DFNB10 locus using 34 novel microsatellite markers, genomic  
structure, and exclusion of six known genes in the region

## TITLE

JOURNAL Genomics 68 (1), 22-29 (2000)  
 MEDLINE 20408883  
 REFERENCE 2 (bases 1 to 57000)  
 AUTHORS Shimizu, N., Kudo, J. and Shibuya, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-FEB-2000) to the DDBJ/EMBL/Genbank databases.  
 Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular  
 Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan  
 (E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-3351-2370,  
 Fax: 81-3-3351-2370)  
 Location/Qualifiers

## FEATURES

source

1. 57000

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_line="FLEB14-14"

/cell\_type="pre-pro-B cell"

/chromosome="21"

/clone\_1lb="Keio BAC library"

/map="21q22.3"

join(957..1081,4085..4237,4997..5208)

/gene="TFP1"

join(997..1081,4085..4237,4997..5013)

/gene="TFP1"

/codon\_start=1

/product="trefol1 factor 1"

/protein\_id="BAB13729.1"

/db\_xref="GI:10280534"

/translation="MATMENKVIKALIVISMLAIGLAEQETETCTVAPRERONCGFP  
GVTSPCANKGCCFDDTVKRVPCFYPNTIDVPEECF"

join(16511..16613,17462..17611,19860..20006,20946..21132)

/gene="TFP2"

join(16511..21132)

/gene="TFP2"

join(16535..16613,17462..17611,19860..20006,20946..20959)

/gene="TFP2"

/codon\_start=1

/product="trefol1 factor 2"

/protein\_id="BAB13730.1"

/db\_xref="GI:10280535"

/translation="MGRDAQLIALILVIGLALAGSEKSPQCSRLSPHNNTNGF  
PGITSDQCPDNGCCFSSVTGVPWCFHPKQSDQCVMEVSRNRMCGRGDISPECA  
SRKCCFNFLEFVPCFPRSVEDCH"

join(52117..52198,53860..54006,55222..55412)

/gene="TFP3"

join(52117..55412)

/gene="TFP3"

join(52138..52198,53860..54006,55222..55235)

/gene="TFP3"

/codon\_start=1

/product="trefol1 factor 3"

/protein\_id="BAB13731.1"

/db\_xref="GI:10280536"

/translation="MGDIVIALISSSSAEERYVLSANQCAVPAKDRVDCGYPVYTRKE  
CNRGCCFDSRLPGVPCFPLDEACTP"

BASE COUNT 13757 a 13698 c 13997 g 15548 t

ORIGIN

Query Match 66.6%; Score 607; DB 9; Length 57000;

Best Local Similarity 100.0%; Pred. No. 1.1e-162;

Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 tgcctgttcttcaatgttactgtacaaagaaacccaggaatagtaagta 71  
 |||||||  
 DB 55824 TGCCTGTTTCTTAATGTCTTACTGTACAAAGAAACCCAGGAATAGTCAAGTA 55765

OY 72 ttgaacagtagcagagtgctgtgaataaagagcaacttggaaagacagttattg 131  
 |||||||  
 DB 55764 TTGAACAGTAGCAGAGTGTGTGAATAAAGGACCACTTTGGAACACAGTTTATTGG 55705

OY 132 ctctgtcttccaccagaagaagcttgatcttltgaaactcttactgaatgtatt 191  
 |||||||  
 DB 55704 CTCTGTCTTCCACCAAGAAAGACTGTGATTTTGAACCTTACTCAAAATGATTTT 55645

OY 192 ttctgtcttcccgagaagcgaacttacagtgcttcttaggtcttcgttgacgtggt 251  
 |||||||  
 DB 55644 TTCTGTCTTCCCGAGAGACGGCAGCTTACAGTGTCTTAGGCTTCTGTACGTGGT 55585

OY 252 gccagcttgatccaataatcccttgatgacgtcagctcagcttccttagggagcttcttc 311  
 |||||||  
 DB 55584 GCCAGTCTGATTCAAATATCTCTTGCATGCACAGCTCTTAGGAGACTTTTCTCTG 55525

OY 312 ccccttaggagcttgagagacttcccttgacacccctccgccccttcccgagcga 371  
 |||||||  
 DB 55524 CCTTGAAGGCTTGAGGAGCTCTCCCTGACACCTCTCCCTCTCCACGAGCGAGCA 55465

OY 372 gaataaagcacaaccccaagaagctcgaagcgaagacgttccctgggtggaagatg 431  
 |||||||  
 DB 55464 GAATTAAGCACAACCTCAGAAAGCTCAGGACGAGAAACTGTCTCGGGTGGAGCATG 55405

OY 432 ggaacttattcgttlaagacataagctcagatatagaacttccagagaagcgcttgc 491  
 |||||||  
 DB 55404 GGAACCTTATTCTGTTAAGACATCAGCTCCAGATATGAACCTTCAGCAGAAAGCGCTTGGC 55345

OY 492 gggagcagaagggagcagaagaagctgagatgacagtgcttgccgaacatccagccggca 551  
 |||||||  
 DB 55344 GGGAGCAAGAGGACAGAAAAGCTGAGATGAACGTGCTGGCAGCAATCCACCGGGCA 55285

OY 552 aggggtctccgaagctcgaatcccccggcgaggcgagctggagggtgctccagaaggtgc 611  
 |||||||  
 DB 55284 AGGGTCTCCGAGGCTTGCACTCCCGCGCGGGGCGAGCTGGAGGTGCTTCAGAAAGTGC 55225

OY 612 attctgc 618  
 |||||||  
 DB 55224 ATTCTGC 55218

## RESULT 2

AP001623

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AP001623 157739 bp DNA PRI 03-JUN-2000  
 Homo sapiens genomic DNA, chromosome 21, clone: KB169B4, MX1-D21s171  
 region, complete sequence.  
 AP001623  
 AP001623.1 GI:7670577  
 HTG.  
 Homo sapiens pre-pro-B cell cell\_line:FLEB14-14 DNA, clone\_1lb:Keio  
 BAC library clone:KB169B4.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 157739)  
 Shimizu, N., Kudo, J. and Shibuya, K.  
 Homo sapiens genomic DNA, chromosome 21, clone: KB169B4, MX1-D21s171  
 region  
 Published Only in DataBase (2000) In press  
 2 (bases 1 to 157739)  
 Shimizu, N., Kudo, J. and Shibuya, K.  
 Direct Submission  
 Submitted (04-APR-2000) to the DDBJ/EMBL/Genbank databases.  
 Nobuyoshi Shimizu, Keio University, School of Medicine, Molecular  
 Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160-8582, Japan  
 (E-mail: nshimizu@med.keio.ac.jp, Tel: 81-3-3351-2370,  
 Fax: 81-3-3351-2370)  
 Location/Qualifiers  
 1. 157739  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="FLEB14-14"  
 /cell\_type="pre-pro-B cell"  
 /chromosome="21"  
 /clone\_1lb="Keio BAC library"

BASE COUNT

41302 a 39955 c 38798 g 37684 t

## ORIGIN

## Query Match

66.6%: Score 607; DB 9; Length 157739;

Best Local Similarity 100.0%: Pred. No. 1.2e-162; Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 12 tgcctgtgttcataatgttactgttacaagaacaaacccaagatagtaagta 71
    |||||||
Db 42041 TGCGTGTGTTCATTAAGTTACTGTACAAAGAAACAAACCCAGGATGTACAACTA 42100

Qy 72 ttgaacagtagcgaagtggtgttgaataaagaacacatttgaagaacagttatttg 131
    |||||||
Db 42101 TTGAACAGTAGCAGAGAGGCTGTGAATAAAGACACCTTTGAAGACAGACTTTATTTGG 42160

Qy 132 ctgtctgtcttcacacaagaagaacttggatatttgaanaacttactactgaatgtattc 191
    |||||||
Db 42161 CTGTCTGTCTTCACCAAAAGACTGTGATTTTGAATAACTTACTGAATGTATTTT 42220

Qy 192 ttctgtcttcacagaagaagcgaactacagtgcttcctagacttctctgtgaagtggt 251
    |||||||
Db 42221 TTTCTGTCTTCACGAGAGGCGCATTACAGTGTCTTCTAGGCTTTCTGTGACGTGGGT 42280

Qy 252 gccagctcgtatcaaaaatccttgcacatgcacatgcacatgcacatgcacatgcacat 311
    |||||||
Db 42281 GCCAGTCTGATCAAAATATCTTGATGCATGCATGCATGCATGCATGCATGCATGCATGC 42340

Qy 312 ccccttgagccttgagcagactcctccctgcacacccctcccgccctccccaagaagcagca 371
    |||||||
Db 42341 CCCCTTGAGGCGCTGGGCGAGACTCTCCCTCGAACCCCTCCGCTCCACAGCAGCAGCA 42400

Qy 372 gaataaagacaacacccaagaagcttcacagcagaagaactgtccctcgtgtgagacatg 431
    |||||||
Db 42401 GAATTAAGCAACACTTCAGAAAGTCTCAGGACGAGAAACTGTCTCGGTTGAGCATG 42460

Qy 432 ggaaccttatctgttaagacalcagcgtccagatatagaacttcaagcagaagcgttgcc 491
    |||||||
Db 42461 GGACCTTATATCTGTTAAGACATCAGGCTCCAGATATGAACTTTCAGCAGAAAGCGCTTCC 42520

Qy 492 gggagcagaagaagcagaagaagcgtgagatgaacagtgtcctgcagcaatcacagccgggca 551
    |||||||
Db 42521 GGGAGCAAAAGGACAGAAAGCTGAGATGAGACAGTGCCTGGCAGCAATCAACAGCGGGCA 42580

Qy 552 agggatgctcagcagcgtcgcaccccgccgagcgggagcagctgagagtgctcctaagaagtgcc 611
    |||||||
Db 42581 AGGGTGCCTCCAGCTCGCATCCCGGGCGGGGCGACACTGGAAGTGCCTCAGAAAGTGC 42640

Qy 612 attctgc 618
    |||||||
Db 42641 ATTCTGC 42647

RESULT 3
AC015555/c AC015555 181511 bp DNA PRI 14-DEC-2000
DEFINITION Homo sapiens chromosome 21 clone RP11-113F1, complete sequence.
AC015555
AC015555.13 GI:11602942
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 181511)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Mao,J., Lam,B., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
Yu,S. and Davis,R.W.
Unpublished
JOURNAL 2 (bases 1 to 181511)
REFERENCE Bruno,D., Conn,L., Dela Rosa,M., Federspiel,N., Foreman,P.,
AUTHORS Glukhov,S., Hansen,N., Hyman,R., Mao,J., Marathe,R.,

```

```

TITLE Morehouse,A.J., Oefner,P., Palm,C.J., Ramirez,D., Wilhelmy,J.,
JOURNAL Yu,S. and Davis,R.W.
Direct Submission
Submitted (17-NOV-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
REFERENCE 3 (bases 1 to 181511)
AUTHORS Abola,A.P., Bruno,D., Conn,L., Dela Rosa,M., Faulkner,D.,
Federspiel,N., Glukhov,S., Hansen,N., Herman,Z.S., Hyman,R.,
Lam,B., Mao,J., Marathe,R., Miranda,M., Morehouse,A.J., Nguyen,M.,
Oefner,P., Palm,C.J., Ramirez,D., Southwick,A.M., Wilhelmy,J.,
Yu,S. and Davis,R.W.
Direct Submission
Submitted (14-DEC-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT On Dec 8, 2000 this sequence version replaced gi:11386278.
----- Genome Center
Center: Stanford DNA Sequencing and Technology Development
Center
Center code: SDSRDC
Web site: http://sequence-www.stanford.edu/group/human/
Contact: hum.info@sequence.stanford.edu
----- Project Information
Center project name: 694
Center clone name: RP11-113F1

----- Summary
This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

Bases 1-126158 of this BAC clone overlap with the complement of
bases 1-126554 of KB169B4 (AP001623) and bases 105422-181511 of
this BAC clone overlap with the complement of bases 73187-149850 of
KB1430A10 (AP001622).
FEATURES
    source
        location/Qualifiers
            1..181511
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="21"
            /clone="RP11-113F1"
            /clone_1lb="RP11 human BAC library 11"
            102209
            /note="Single-stranded, single chemistry with consensus
            quality below 30."
            102988..102990
            /note="Single-stranded, single chemistry with consensus
            below 30."
            125544
            /note="Single-stranded, single chemistry with consensus
            quality below 30."
            125566
            /note="Single-stranded, single chemistry with consensus
            quality below 30."
            166246..166247
            /note="Single-stranded, single chemistry with consensus
            quality below 30."
    misc_feature
        46156 c 45935 g 45584 t
BASE COUNT 43836 a 46156 c 45935 g 45584 t
ORIGIN
Query Match 66.6%: Score 607; DB 9; Length 181511;
Best Local Similarity 100.0%: Pred. No. 1.2e-162;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 12 tgcctgtgttcataatgttactgttacaagaacaaacccaagatagtaagta 71
    |||||||
Db 84525 TGCGTGTGTTCATTAAGTTACTGTACAAAGAAACAAACCCAGGATGTACAACTA 84466

```

QY 72 ctgacacgtacgagagtggtgtgaaataaaggacacttctggaagacagtttattcgtg 131  
 |||||||  
 Db 84465 TTGAACAGTAGCAGAGTGGTTGTGAATAAAGAGCACCTTTGGAAGACAGTTTATTGG 84406  
 |||||||  
 QY 132 ctgtgtctcttaccagaagaactgtgtatttctgaaacttctactgaagaatctt 191  
 |||||||  
 Db 84405 CTGTGCTCTCTTCACCAAGAAAGACTTGTGATTTTGAATAACTTCTACCTGAATGTATTT 84346  
 |||||||  
 QY 192 ttctgtcttcccgaggagacgagacttacagtgcttcttaggcttctgtgaagctggt 251  
 |||||||  
 Db 84345 TTGTGTGCTTCCGAGAGACGGCAGCTTACAGTCTTCTAGGCTTCTCTGTGAGCTGGCT 84286  
 |||||||  
 QY 252 gccagctctgacatcaaatcctctgcatgacagctgacgtcctccttaggagctcttctg 311  
 |||||||  
 Db 84285 GCCAGCTGTGATTCAAATATCTCTGATGACACTGACAGCTCTTTAGGAGAGCTTTTCTGT 84226  
 |||||||  
 QY 312 ccccttagagctgggagagactctccctctgacacctccgacctcccaagagagagca 371  
 |||||||  
 Db 84225 CCCTTGAAGCCTGGGACAGACTCTCCCTGACACCTCCCGCTCCACGACGACGCA 84166  
 |||||||  
 QY 372 gaataaagacaacactcagaagaagctcaggaacagcaagaaactgtcctcggtgagcatg 431  
 |||||||  
 Db 84165 GAATTAAGCACAACCTCAGAAAGTCTCAGGACGAGAAAGTGTCTCGGTGAGCATG 84106  
 |||||||  
 QY 432 ggaccttattcgttaagacatcaggtcctcagatatgaacttcaagagaagcgcttgc 491  
 |||||||  
 Db 84105 GGACCTTTATTCCTTAAGACATCAGGCTCCAGATATGAACTTTCAGCAGAAAGCCTTGCC 84046  
 |||||||  
 QY 492 gggagcaaaaggagagaagaactgagatgaacagtgctgagcaatcaacagcgagca 551  
 |||||||  
 Db 84045 GGGGCGCAAGAGGACGAAAGAGTGAATGAACATGCTGCGCAGCAATCACAGCGGGGCA 83986  
 |||||||  
 QY 552 aggggtcctcagagcctcgcattcccccggcgagggcgagctggaagtgcctcagaagtgc 611  
 |||||||  
 Db 83985 AGGGGTCTCGAGAGCTGTGCAATCCCGCGGGGCGAGCTGGAAGTCTTCAGAGAGTGC 83926  
 |||||||  
 QY 612 attctcgc 618  
 |||||||  
 Db 83925 ATTCTGC 83919

RESULT 4  
 AP001746  
 LOCUS AP001746 340000 bp DNA PRI 30-MAY-2000  
 DEFINITION Homo sapiens genomic DNA, chromosome 21q, section 90/105.  
 ACCESSION AP001746 AL163291 BA000005  
 VERSION AP001746.1 GI:7768741  
 KEYWORDS  
 SOURCE Homo sapiens DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (sites)  
 Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,  
 Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,  
 Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,  
 Polley,A., Menzel,U., Delabar,U., Kumpf,K., Lehmann,R.,  
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,  
 The DNA sequence of human chromosome 21. The chromosome 21 mapping  
 and sequencing consortium  
 Nature 405 (6784), 311-319 (2000)  
 2 (bases 1 to 340000)  
 20289799  
 JOURNAL Hattori,M., Fujiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,  
 MEDLINE Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,  
 REFERENCE Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,  
 AUTHORS Polley,A., Menzel,U., Delabar,U., Kumpf,K., Lehmann,R.,  
 Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,  
 Zimmerman,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,  
 Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyama,S.,  
 Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstiek,G.,  
 Hornischer,K., Barandt,P., Schafte,M., Schoen,O., Desario,A.,

TITLE Reichelt,J., Kauer,G., Bloecker,H., Ramser,J., Beck,A., Klages,S.,  
 Henny,S., Rieseemann,L., Dagand,E., Weinmeyer,S., Borzym,K.,  
 Gardiner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and  
 Vaspou,M.L.  
 JOURNAL Direct Submission  
 Submitted (10-APR-2000) to the DDBJ/EMBL/GenBank databases. The  
 Chromosome 21 Mapping and Sequencing Consortium: \* RIKEN Genomic  
 Sciences Center, Human Genome Research Group \* Institute of  
 Molecular Biotechnology, Genome Analysis \* Keio University School  
 of Medicine, Dept. of Molecular Biology \* GBF, Dept. of Genome  
 Analysis \* Max-Planck Institute for Molecular Genetics (addresses  
 see below)  
 On May 30, 2000 this sequence version replaced gi:7717398.  
 The chromosome 21 mapping and sequencing consortium consisting of  
 \* RIKEN Genomic Sciences Center, Human Genome Research Group, \*  
 Sagami-hara 228-8555, Japan,  
 \* e.mail: hattori@gs.riken.go.jp  
 \* URL: http://hgp.gsc.riken.go.jp/  
 \* Institute of Molecular Biotechnology, Genome Analysis, \*  
 Beutenbergstrasse 11, D-07745 Jena, Germany,  
 \* e.mail: gscj-submit@genome.imb-jena.de  
 \* URL: http://genome.imb-jena.de/  
 and  
 \* Keio University School of Medicine, Molecular Biology, \* Tokyo  
 160-8582, Japan,  
 \* e.mail: nshimizudmb-med.keio.ac.jp  
 \* URL: http://www.dmb.med.keio.ac.jp/  
 and  
 \* GBF, Dept. of Genome Analysis,  
 \* Mascheroder Weg 1, D-38124 Braunschweig, Germany, \* e.mail:  
 info.genome@gbf.de  
 \* URL: http://genome.gbf.de/  
 and  
 \* Max-Planck Institute for Molecular Genetics,  
 \* Inestrasse 73, D-14195 Berlin, Germany,  
 \* e.mail: info-chr21@molgen.mpg.de  
 \* URL: http://chr21.fz-berlin.mpg.de/  
 AL163291: Submitted (10-APR-2000).  
 Location/Qualifiers  
 1. 340000  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 <1..44256  
 /note="Accession No. AP001621"  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /map="21q22.3"  
 /clone="KB1342D7, 5' partial"  
 /clone.lib="Keio BAC library"  
 <1..124582  
 /note="Accession No. AP001622"  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="21"  
 /map="21q22.3"  
 /clone="KB1430A10, 5' partial"  
 /clone.lib="Keio BAC library"  
 join(<310..374,9668..9804)  
 /gene="PREDA6"  
 /note="exon prediction only"  
 310..9804  
 /gene="PREDA6"  
 310..374  
 /gene="PREDA6"  
 /number=1  
 complement(626..705)  
 /note="LIMC4"  
 /rpt\_family="LINE/L1"  
 /rpt\_type="DISPERSED  
 706..1048



[illegible]



```
|||||
Db 341 ACAGGCCACGTAC-TCTCAGCAGACTGGAGACAGCAAGCAGACGACCCAGC 283
QY 866 atgagagcgtctgagcagcatgacacgctggctccggagcgc 911
|||||
Db 282 ATGAGAGCGCTGTGGACGATGACACCGTGGCTCCGGAGCGC 237

RESULT 7
AX193486/c AX193486 480 bp DNA PAT 15-AUG-2001
LOCUS Sequence 1053 from Patent WO0149716.
ACCESSION AX193486
VERSION AX193486.1 GI:15211437
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 480)
XU,J., Lodes,M.J., Secrist,H., Benson,D.R., Meagher,M.J.,
AUTHORS Stolk,J.A., King,G.E., Wang,T. and Jiang,Y.
TITLE Compounds for immunotherapy and diagnosis of colon cancer and
methods for their use
JOURNAL Patent: WO 0149716-A 1053 12-JUL-2001;
CORIXA CORPORATION (US)
FEATURES
source location/Qualifiers
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 78 a 154 c 141 g 107 t
ORIGIN
Query Match 44.2%; Score 403; DB 6; Length 480;
Best Local Similarity 91.2%; Pred. No. 2.5e-104;
Matches 468; Conservative 0; Mismatches 0; Indels 45; Gaps 2;

QY 399 caggcagaagaactgtctcgtgggtgagccttgacttattcgttaagacatcagc 458
|||||
Db 480 CAGGACGAAAGACTGTCTCGGTGGAGCATGGGACCTTATTCGTTAAGACATCAGC 421
459 tccgatatgaacttccagcagaagcgctgcccggagcaaaaggagcaagaactgaga 518
420 TCCAGATATGAACTTTCAGCAGAGCGCTTGCCGGAGCAAGAGCAGAAAGCTGAGA 361
QY 519 tgaacagtgcctgagcagaatcacagccgggcaagggtgctcgagctcgatccccg 578
|||||
Db 360 TGAACAGTGCCTGGCAGCAATCACAGCCGGGCAAGGGGTGCTCCAGCCTCCATCCCCG 301
QY 579 gccgggggagcgtggaagtgctcagaaggtgcatctcgtctctgaggggcttgaac 638
|||||
Db 300 GCCGGGGGACACTGGAGGTGCTCAGAAAGGTGCAATTCGCTTCGCAAGGGGCTTGAAC 241
639 accaaggaactccaggagatcctgaggtcaaaagcagccccgggtgtgtgacctctg 698
|||||
Db 240 ACCAAGGAGATCCAGGGATCTCTGAGTCAAAAGCAGAGCCCGGTTGTGCACTCTGG 181
QY 699 ggggtgacatggggtgagcgcagtcacacctgtctctggtgagcagggacactgttg 758
|||||
Db 180 GGGTGAATGGGGGTAGCGGAGTCCACCCCTGCTTGCTGGCAGCGCACATGCTGTTG 121
QY 759 caggtgtcccaagaagacctgtcagctgcagaagccctgtcgtggaagaagcccaagta 818
|||||
Db 120 CA-----GACAGGCCACAGTA 105
QY 819 ctctctcagcagagctgagagacagcagaagccagaccagccatgagagcgctc 878
|||||
Db 104 C-TCTCTAGCAGAGCTGGAGGACAGCAAGGCCAGAGACCCAGCATGAGAGCGGTC 46
QY 879 tggcagcattgacacacgtgtggtctccgggagcgc 911
|||||
Db 45 TGGCAGCGCATGACCAACGCTGGGCTCCGGGAGCGC 13
```

```
RESULT 8
HUMPIBX/c HUMPIBX 480 bp mRNA PRI 07-JAN-1995
LOCUS Human secretory protein (PI.B) mRNA, complete cds.
DEFINITION LI5203
ACCESSION LI5203.1 GI:402482
VERSION
KEYWORDS secretory protein.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 480)
Hauser,F., Poulsom,R., Chinery,R., Rogers,L.A., Hanby,A.M.,
AUTHORS Wright,N.A. and Hoffmann,W.
TITLE hpl.B, a human P-domain peptide homologous with rat intestinal
trefoil factor, is expressed also in the ulcer-associated cell
lineage and the uterus
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 90 (15), 6961-6965 (1993)
MEDLINE 93348192
FEATURES
source location/Qualifiers
1..480
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="Intestine, stomach, uterus"
mRNA 1..460
/gene="PI.B"
/note="minor transcript"
mRNA 1..480
/gene="PI.B"
/note="major transcript"
1..480
/gene="PI.B"
36..98
/gene="PI.B"
36..278
/gene="PI.B"
/codon_start=1
/product="secretory protein"
/protein_id="AA59981.1"
/db_xref="GI:402483"
/translation="MAARALCMGLVLTALSSSAEYVGLSANQCAVPAKRDVDCGY
PHYTPKRNRRGCGFDSIRIPGVWCFPIDEAECT"
437..442
polyA_signal /gene="PI.B"
BASE COUNT 78 a 154 c 141 g 107 t
ORIGIN
Query Match 44.2%; Score 403; DB 9; Length 480;
Best Local Similarity 91.2%; Pred. No. 2.5e-104;
Matches 468; Conservative 0; Mismatches 0; Indels 45; Gaps 2;
```



AXI06398/c	AXI06398	250 bp	DNA	PAT	30-APR-2001
LOCUS	Sequence	179	from Patent WO0125272.		
DEFINITION	AXI06398				
ACCESSION	AXI06398.1	GI:13922077			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM	human.				
	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	1 (bases 1 to 250)				
	Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A.				
JOURNAL	Compositions and methods for therapy and diagnosis of prostate				
	cancer				
	Patent: WO 0125272-A 1/19 12-APR-2001;				
FEATURES	CORIXA CORPORATION (US)				
source	Location/Qualifiers				
	1..250				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	55 a	75 c	63 g	57 t	
ORIGIN					

Query Match	25.9%	Score 236.2	DB 6	Length 250
Best Local Similarity	98.8%	Pred. No. 1.4e-56		
Matches 238	Conservative 0	Mismatches 5	Indels 0	Gaps 0
Qy 413	tgctccgcggtgagagcatgtgacattatctgttaagaacatcaggtccagatcgaact	472		
Db 241	TTTTTTTCGGGGTGGAGCATAGGACCTTATTCCTTAAACATCAGGCTCCAGATATGAACT	182		
Qy 473	ttcagcagaagcgcgtctgcgggaggaacaaaggacagaaaaagctgagatgaacagtccttg	532		
Db 181	TTTCAGCAGAAACGCGCTTCCGCCGGAGCAAAAGGACAGAAAAGCTGAGATGAAACAGTCCCTGG	122		
Qy 533	cagaatcacagccggggagaaagggtgtctcgagccctcgatccccccggccgggggagactg	592		
Db 121	CAGCATATACAGCCGGGGAAGGGTCTCCGAGCCCTCGCATCCCGCGGGGAGCTGG	62		
Qy 593	gaggtgctccagaaggtgcatctgtctctcctcagagggttgaacacccaaggacctca	652		
Db 61	GAGGCGCTTCAGAAAGTGCATTTCTCTTCTTCACAGGGCTTGAACACCAAGCAAGCACTCCA	2		
Qy 653	g 653			
Db 1	g 1			
RESULT 12				
AX140689/c				
LOCUS	AX140689	250 bp	DNA	
DEFINITION	Sequence 179 from patent WO0134802.		PAT	31-MAY-2001
ACCESSION	AX140689			
VERSION	AX140689.1	GI:14280800		
KEYWORDS	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 250)			
JOURNAL	Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,			
FEATURES	Reed,S.G., Kalos,M.D., Retter,M.W., Stolk,J.A., Day,C.H.,			
SOURCE	Skelky,Y.A. and Wang,A.			
BASE COUNT	Compositions and methods for the therapy and diagnosis of prostate			
	Cancer			
	Patent: WO 0134802-A 179 17-MAY-2001;			
	CORIXA CORPORATION (US)			
	Location/Qualifiers			
	1..250			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	55 a 75 c 63 g 57 t			

ORIGIN	Query Match	25.9%	Score 236.2	DB 6	Length 250	
	Best Local Similarity	98.8%	Pred. No. 1.4e-56			
	Matches 238	Conservative 0	Mismatches 3	Indels 0	Gaps 0	
QY	413	tytctctcggtggaagcaatcggagaccttattctgttaagaacatcagagcttcagatatgaact	472			
Db	241	TTTTTTTCGGGGTGGAGCAATGGGACCTTTATTTGCTTAAAGCATCAGGCTCCAGATATGAAGT	182			
QY	473	ttcagcagaagcgccttgcgcggagagcaaaagggacagaaagctgagatataacagtgctctgg	532			
Db	181	TTTCAGCAGAAAGCGCTTCCCGGAGCAAAAGGGACAGAAAAGCTGAGATGTAACAGTGCCTGG	122			
QY	533	cagcaatcaacagccgggccaaggggtgctctcgagccttcgcatcccccggcccggggcagctg	592			
Db	121	CAGCAATCACAAGCCGGGCAAGGGTGTCTCCGAGCCTCGCATCCCCCGGGCGGCGAGCTG	62			
QY	593	gagtgctctcagaaggtgcatcttgctcttcctcgcaggggcttgaaacaccacaaggaactcca	652			
Db	61	GAGTGCTCTCAGAAAGTGTCATTCTCTTCTCTGACAGGGCTGTAAACACCCNAAGGCACCTCCA	2			

QY	653 g	653
Db	1 g	1
RESULT	13	
LOCUS	AX106296/c	
DEFINITION	Sequence 77 from Patent WO0125272.	PAT
ACCESSION	AX106296	
VERSION	AX106296.1	
KEYWORDS	GI:13921982	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
ATTNOS	1 (bases 1 to 248)	
TITLE	Xu,J., Skeiky,Y.A., Reed,S.G. and Cheever,M.A. Compositions and methods for therapy and diagnosis of prostate cancer	
JOURNAL	Patent: WO 0125272-A 77 12-APR-2001; CORIXA CORPORATION (US)	
FEATURES	Location/Qualifiers	
source	1..248	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
BASE COUNT	55 a 73 c 63 g 57 t	
ORIGIN		
Query Match	24.2%; Score 220.6; DB 6; Length 248;	
Best Local Similarity	97.5%; Pred. NO. 4.3e-52;	
Matches 235; Conservative	0; Mismatches 4; Indels 2; Gaps 1;	
QY	413	472
Db	239	180
QY	473	532
Db	179	120
QY	533	592
Db	119	62
QY	593	652
Db	61	2



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 07:10:03 ; Search time 519.52 Seconds  
(without alignments)  
1503.356 Million cell updates/sec

Title: US-09-867-034-1

Perfect score: 911  
Sequence: 1 ttttttttttcgtctgttt.....cacctgtggtctcggagcgc 911

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_1101:\*  
1: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
2: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	521.8	57.3	653	AAH34958	Human colon cancer
2	516.8	56.7	834	AAZ33462	Human prostate can
3	516.8	56.7	834	AAZ33629	Human breast tumou
4	403	44.2	480	AAI29499	C886P determined c
5	378.8	41.6	494	AAI29499	Human breast and c
6	305.4	33.5	570	AAI29499	Human colon specif
7	305.4	33.5	570	AAI29499	Polymerase chain seq
8	265.4	29.1	403	AAO28363	Human intestinal t
9	265.4	29.1	403	AAI28039	Human intestinal t
10	265.4	29.1	403	AAI28039	Human intestinal t
11	259.4	28.5	263	AAI22480	Human gene signatu

12	254.8	28.0	594	22	AAH34957	Human colon cancer
13	236.2	25.9	250	19	AAV61325	Extended cDNA sequ
14	236.2	25.9	250	19	AAV58649	Prostate tumour sp
15	236.2	25.9	250	21	AAH06412	Human immunogenic
16	236.2	25.9	250	22	AAH06412	Human immunogenic
17	236.2	25.9	250	22	AAH06412	Human immunogenic
18	236.2	25.9	250	22	AAH06412	Human immunogenic
19	236.2	25.9	250	22	AAH06412	Human immunogenic
20	236.2	25.9	250	22	AAH06412	Human immunogenic
21	236.2	25.9	250	22	AAH06412	Human immunogenic
22	236.2	25.9	250	22	AAH06412	Human immunogenic
23	236.2	25.9	250	22	AAH06412	Human immunogenic
24	236.2	25.9	250	22	AAH06412	Human immunogenic
25	236.2	25.9	250	22	AAH06412	Human immunogenic
26	236.2	25.9	250	22	AAH06412	Human immunogenic
27	236.2	25.9	250	22	AAH06412	Human immunogenic
28	236.2	25.9	250	22	AAH06412	Human immunogenic
29	236.2	25.9	250	22	AAH06412	Human immunogenic
30	236.2	25.9	250	22	AAH06412	Human immunogenic
31	236.2	25.9	250	22	AAH06412	Human immunogenic
32	236.2	25.9	250	22	AAH06412	Human immunogenic
33	236.2	25.9	250	22	AAH06412	Human immunogenic
34	236.2	25.9	250	22	AAH06412	Human immunogenic
35	236.2	25.9	250	22	AAH06412	Human immunogenic
36	236.2	25.9	250	22	AAH06412	Human immunogenic
37	236.2	25.9	250	22	AAH06412	Human immunogenic
38	236.2	25.9	250	22	AAH06412	Human immunogenic
39	236.2	25.9	250	22	AAH06412	Human immunogenic
40	236.2	25.9	250	22	AAH06412	Human immunogenic
41	236.2	25.9	250	22	AAH06412	Human immunogenic
42	236.2	25.9	250	22	AAH06412	Human immunogenic
43	236.2	25.9	250	22	AAH06412	Human immunogenic
44	236.2	25.9	250	22	AAH06412	Human immunogenic
45	236.2	25.9	250	22	AAH06412	Human immunogenic

#### ALIGNMENTS

RESULT 1	AAH34958/c	AAH34958 standard; cDNA: 653 BP.
ID	AAH34958	
AC	AAH34958	
DT	03-SEP-2001	(first entry)
DE	Human colon cancer antigen encoding cDNA seq ID NO:2040.	
XX	Human: colon cancer; colon cancer antigen; diagnosis; detection;	
KW	colorectal carcinoma; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	MO200122920-A2.	
XX		
PD	05-APR-2001.	
XX		
PF	28-SEP-2000; 2000WO-US26524.	
XX		
PR	29-SEP-1999; 99US-0157137.	
XX		
PR	03-NOV-1999; 99US-0163280.	
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Ruben SM, Barash SC, Birse CE, Rosen CA;	
XX		
DR	WPI: 2001-235357/24.	
XX		
DR	P-PSDB: AAG75553.	
XX		
PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,	
XX	useful for preventing, diagnosing and/or treating colorectal cancers -	

PS Claim 1: Page 3533; 9803bp; English.  
 XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patient's own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated PS,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAB77789 represent sequences used in the exemplification of the  
 CC present invention.  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.

SO Sequence 653 BP; 111 A; 193 C; 198 G; 145 T; 6 other;

Query Match 57.3%; Score 521.8; DB 22; Length 653;  
 Best Local Similarity 92.3%; Pred. No. 1.7e-148;  
 Matches 596; Conservative 3; Mismatches 1; Indels 46; Gaps 3;

QY 266 aaatatccttgatgacgacgagcctccttaaggagcttcttcctgacctgagccttg 325  
 DB 653 AAATATCCTTGATGACACGACCTCTTAAGGAGTCTTCTCCCTTAAGGCTT-6 595  
 QY 326 gcagactcctccctgacacccctccctcctccacagacgacgacgaataaagacaa 385  
 DB 594 GCAGACTCTCCCTGACACCTCTCCGCTCCCTCCAGACGACGACCAATAAAGCAAA 535  
 QY 386 cctcaagaagctcagacgacgaagaactgtcctgggtgagacatgagacctatcgt 445  
 DB 534 CCTCAGAAATCTCAGGACGAGAAAGACTGCTCCGCTGGAGACTGGACCTTATTCGT 475  
 QY 446 taagacacagctcagacgacgacgacgacgacgacgacgacgacgacgacgacgac 505  
 DB 474 TAAGACATCAGGCTCCGATATGACATTTCAAGAAAGCTTCCCGGAGCAAAAGGAC 415  
 QY 506 agaaagctgagatgaacagctgcctgagacgaatcacacagccgaggaaggtgtcctgac 565  
 DB 414 AGAAAGCTGAGATGAAACAGTGCCTGGCAGCAATCACAGCCGGGCAAGGCTCCGAGC 355  
 QY 566 ctgcgaccccgagcgaggggagcagctgagaggtgcctcaagaagtgaatctgtcctgc 625  
 DB 354 CTGCGAATCCCGCCGCGGCGGAGCTGAGAGTGCCTCAGAAAGGTGCAATTCGTTCTCTGC 295  
 QY 626 aggggcttgaagaacagacgacgacgacgacgacgacgacgacgacgacgacgacgac 685  
 DB 294 AGGGGCTTGAACCAACAGGACACTCCAGGATCTTGAAGTCAAAAGCAGACGCCCGGTTG 235  
 QY 686 ttgcactccttgagggtgtaaatgaggttagccgagctccacccctgtcctgagctgagc 745  
 DB 234 TTGCACATCCCTGGGGGTGATAGGAGGTACCGCAGTCCACCCCTGTCTTGGCTGGGACG 175  
 QY 746 gcaactggtttgacagctgtccagacaaagccctgtcagctgcagagccctgtcgtg 805  
 DB 174 GCACACTGCTTGTCA-----G 159  
 QY 806 acagggcccaagctacttctcctagcagacgctgagagacgacgaagccagacccagc 865  
 DB 158 ACAGGCCCAAGCTAC-TCCTTAGCAGAGACTGTGAGACAGCAAGGCCAGACCCAGC 100  
 QY 866 atcagagcgctctgagcagcagacgacgacgctgagctcggagcgc 911  
 DB 99 ATCAGAGGCTCTGTGAGGACCATGACACCGTGGGCTCCGGGAGCG 54

# RESULT 2

AAZ33462/c  
 ID AAZ33462 standard; cDNA; 834 BP.

XX  
 AC AAZ33462;

XX  
 DT 08-DEC-1999 (first entry)

XX  
 DE Human prostate cancer-associated EST 40.

XX  
 KW Expressed sequence tag; EST; prostate tumor; antitumor; treatment;  
 KW gene therapy; tissue specificity human; ss.

XX  
 OS Homo sapiens.

XX  
 PN DE19811193-A1.

XX  
 PD 16-SEP-1999.

XX  
 PF 10-MAR-1998; 98DE-1011193.

XX  
 PR 10-MAR-1998; 98DE-1011193.

XX  
 PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX  
 PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A;

XX  
 DR WPT; 1999-519628/44.

XX  
 DR P-PSDB; AAY48258.

PT  
 PT New nucleic acid expressed at high level in prostatic tumor tissue and  
 PT encoded polypeptides, useful for treating cancer and screening for  
 PT therapeutic agents

PS  
 PS Claim 3; 101; 166pp; German.

XX  
 CC This invention describes novel nucleic acid sequences (A) that are  
 CC expressed at high level in prostatic tumor tissue and encode gene  
 CC products or their fragments. The products of the invention have  
 CC antitumor activity. Polypeptides (I) encoded by (A) are used: (i) for  
 CC identifying agents for treatment of prostatic cancer and (ii) for  
 CC therapy of prostate cancer, optionally where expressed by gene therapy  
 CC methods. (A) is also used to isolate full-length genes (for gene therapy)  
 CC and for recombinant production of (I), which can be used to raise  
 CC specific antibodies. (A) are identified by assembly of ESTs (expressed  
 CC sequence tags) before they are analyzed for expression pattern (tissue  
 CC specificity). This approach eliminates many of the false results, as  
 CC regards tissue specificity, associated with known methods that use  
 CC single (usually short) ESTs. AAZ33423-233476 represent expressed  
 CC sequence tags described in the method of the invention.

SO Sequence 834 BP; 173 A; 241 C; 249 G; 171 T; 0 other;

Query Match 56.7%; Score 516.8; DB 20; Length 834;  
 Best Local Similarity 92.3%; Pred. No. 6.7e-147;  
 Matches 596; Conservative 0; Mismatches 2; Indels 48; Gaps 3;

QY 266 aaatatccttgatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 325  
 DB 834 AAATATCCTTGATGACACGACCTCTTAAGGAGTCTTCTCCCTTAAGGCTT-6 595  
 QY 326 gcagactcctccctgacacccctccctcctccacagacgacgacgaataaagacaa 385  
 DB 774 GCAGACTCTCCCTGACACCTCTCCGCTGT---CCAGACGACGACGAAATAAAGCAAA 718  
 QY 386 cctcaagaagctcagacgacgaagaactgtcctcgggttggagatgagacctatcgt 445  
 DB 717 CTCAGAAATCTCAGGACGAGAAAGACTGCTCCGCTGGAGCATGAGACTTATTCGT 658  
 QY 446 taagacacagctcagacgacgacgacgacgacgacgacgacgacgacgacgacgac 505



Db 657 TAAGACATCAGGCTCCAGATATGAACCTTTCAGCAGAGACCGCTTGCCGGGAGCAAGGAGC 598  
QY 506 agaaagcttgatgaaacagtgccctgagcaatcacagccgggaaaggtgtctccagc 565  
Db 597 AGAAAAGCTGAGATGAAACAGTGCCTGGCAGCAATCACAGCCGGGCAAGGGTGTCTCCGAGC 538  
QY 566 ctgcgcatcccccggccgggggagcagctggaagggtgcctcagaaagtgatcttgcctctgc 625  
Db 537 CTCGCATCCCCCGGGGGGAGCTGGAGGTGCCTCAGAAAGGTGATTCGCTTCCCTCCGAC 478  
QY 626 agggagcttgaacacacaaagcagctccagagatcctggaagtaaaagcagagcccggttg 685  
Db 477 AGGGGCTTTGAAACACCAAGGCACTCCAGGGATCTCGAGTCAAAAGCACAGCCCCGGGTGG 418  
QY 686 ttgacatccttgagggtgacatgagggtgagcagctccacccctgtctcttgcttgagcag 745  
Db 417 TTGCACATCTTGGGGGTGACATGGGGGTAGCCGACAGTCCACCTGTCTTGGTGGCAGC 358  
QY 746 gcacactgtgttgagctgtgccagacaaagccctgtcagctgcagagcccttgctgg 805  
Db 357 GCACACTGTGTTGCA-----G 342  
QY 806 acagagcccaagtaactctctcagcagagctggaagacagcaagccagagaccagccagc 865  
Db 341 ACAGGCCCAAGCTAC-TCTTCAGCAGAGCTGGAGGACACCAAGGCCAGGACCCCGCAGC 283  
QY 866 atgcagagcagctctgcagcagcatcacacagctgggctccggagcagc 911  
Db 282 ATGCAGAGCGCTCTGGCAGCATGACACCGTGGGCTCCGGGACGC 237

RESULT 3  
AA233629/c  
ID AA233629 standard; cDNA; 834 BP.

AC AA233629;  
DT 08-DEC-1999 (first entry)  
XX Human breast tumour-associated EST 19.  
DE  
XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;  
KW treatment; tumour; cytostatic; medicament; ss.  
XX  
OS Homo sapiens.  
XX  
PN DE19813839-A1.  
XX  
PD 23-SEP-1999.  
XX  
PE 20-MAR-1998; 98DE-1013839.  
XX  
PR 20-MAR-1998; 98DE-1013839.  
XX  
PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosenthal A.  
XX WPI; 1999-528981/45.  
DR P-PSDB; AAY48551.  
XX  
PT Human nucleic acid sequences and protein products from tumor breast  
XX tissue, useful for breast cancer therapy -  
XX  
PS Claim 3; 99; 188pp; German.  
XX  
CC This invention describes novel human nucleic acid sequences from tumor  
CC breast tissue which have cytostatic activity. The nucleic acid sequences  
CC can be used to produce and isolate full-length gene sequences. They can  
CC be used to express proteins, which can be used as tools to find an  
CC activity against breast cancer. The sequences can be used in sense or  
CC antisense form. They are especially useful for medicaments for gene  
CC therapy to treat breast cancer. AA233611-248617 represents expressed

CC sequence tags described in the method of the invention.  
XX  
SQ Sequence 834 BP; 173 A; 241 C; 249 G; 171 T; 0 other;

Query Match 56.7%; Score 516.8; DB 20; Length 834;  
Best Local Similarity 92.3%; Pred. No. 6,7e-147;  
Matches 596; Conservative 0; Mismatches 2; Indels 48; Gaps 3;

QY 266 aaatatccttgatgacatgacagctccttaggagatcttctgccttgagccttg 325  
Db 834 AAAATATCCTTGATGACACTCAGCTCCTTAGGGAGTCTTTCTGCGCTTGAGGCTTGG 775  
QY 326 gcacatcctccctgaaacccctccgcccctccacagcagcagaaataaagacaa 385  
Db 774 GCAGACTTCCCCCTGACACCTCCCGCTGT---CCACAGCGACAGCAAAATTAACACAA 718  
QY 386 cctcagaagctcagcagcagaagaaactgtcctcgggtgagacatggaaccttattcgt 445  
Db 717 CCTCAGAAAGTCTCAGCGACGAAAGTCTCGGGTGGAGCATGGGACCTTATTGCT 658  
QY 446 taagacatcaggtctccagatataatgaacttcagcagaagcgttcgagagcaaaaggac 505  
Db 657 TAAGACATCAGGCTCCAGATATGAACCTTTCAGCAGAAAGCGCTTGCCGGAGCAAGGAGC 598  
QY 506 agaaagcttgatgaaacagtgccctgagcaatcacagccgggaaaggtgtctccagc 565  
Db 597 AGAAAAGCTGAGATGAAACAGTGCCTGGCAGCAATCACAGCCGGGCAAGGTGTCTCCGAGC 538  
QY 566 ctgcgcatcccccggccgggggagcagctggaagggtgcctcagaaagtgatcttgcctctgc 625  
Db 537 CTCGCATCCCCCGGGGGGAGCTGGAGGTGCCTCAGAAAGGTGATTCGCTTCCCTGC 478  
QY 626 agggagcttgaacacacaaagcagctccagagatcctggaagtaaaagcagagcccggttg 685  
Db 477 AGGGCTTTGAAACACCAAGGCACTCCAGGATCTCGAGTCAAAACACACACCCCGGTGG 418  
QY 686 ttgacatccttgagggtgacatgagggtgagcagcagctccacccctgtcttgcttgagcag 745  
Db 417 TTGCATCTCTTGGGGGTGACATGGGGGTAGCCGACAGTCCACCTGTCTTGGTGGCAGC 358  
QY 746 gcacactgtgttgagctgtgccagacaaagccctgtcagctgcagagcccttgctgg 805  
Db 357 GCACACTGTGTTGCA-----G 342  
QY 806 acagagcccaagtaactctcagcagagctggaagacagcaagccagagaccagccagc 865  
Db 341 ACAGGCCCAAGCTAC-TCTTCAGCAGAGCTGGAGGACACCAAGGCCAGGACCCCGCAGC 283  
QY 866 atgcagagcagctctgcagcagcatcacacagctgggctccggagcagc 911  
Db 282 ATGCAGAGCGCTCTGGCAGCATGACACCGTGGGCTCCGGGACGC 237

RESULT 4  
AA129499/c  
ID AA129499 standard; cDNA; 480 BP.

AC AA129499;  
DT 12-OCT-2001 (first entry)  
XX C886P determined cDNA sequence.  
DE  
XX Human; immunotherapy; diagnosis; colon cancer; colon tumour;  
KW immunogenic; gene therapy; vaccine; colonic cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200149716-A2.  
XX  
PD 12-JUL-2001.  
XX

PF 29-DEC-2000; 2000MO-US355596.  
 XX 30-DEC-1999; 9905-0476296.  
 PR 10-JAN-2000; 2000US-0480321.  
 PR 15-FEB-2000; 2000US-0504629.  
 PR 06-MAR-2000; 2000US-0519444.  
 PR 19-MAY-2000; 2000US-0575251.  
 PR 29-JUN-2000; 2000US-0609448.  
 PR 28-AUG-2000; 2000US-0649811.  
 XX  
 PA (CORI-) CORIXA CORP.  
 PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk JA;  
 PI King GE, Wang T, Jiang Y;  
 DR WPI; 2001-441847/47.  
 XX  
 PT Colon tumor associated proteins and nucleic acids useful for the  
 PT prevention, diagnosis and treatment of colonic cancer -  
 XX  
 PS Claim 2; Page 423; 472pp; English.  
 XX  
 CC The present invention describes colon tumour associated proteins (I) and  
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.  
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and  
 CC (II) may be used in the prevention, diagnosis and treatment of diseases  
 CC associated with inappropriate colon tumour associated protein (TCAP)  
 CC expression, such as colonic cancer. For example, (I) and (II) may be  
 CC used to treat disorders associated with decreased expression by  
 CC rectifying mutations or deletions in a patient's genome that affect the  
 CC activity of TCAPs by expressing inactive proteins or to supplement the  
 CC patient's own production of them. Additionally, (II) may be used to  
 CC produce the TCAP proteins, by inserting the nucleic acids into a host  
 CC cell culturing the cell to express the protein. (II) and its  
 CC complementary sequences may also be used as DNA probes in diagnostic  
 CC polymerase chain reaction (PCR) and hybridisation assays to detect and  
 CC quantitate the presence of similar nucleic acids in samples, and  
 CC therefore which patients may be in need of restorative therapy. (I) may  
 CC also be used as antigens in the production of antibodies against TCAPs  
 CC and in assays to identify modulators of TCAP expression and activity.  
 CC Anti-(I) antibodies and antagonists may also be used to down regulate  
 CC TCAP expression and activity. The anti-(I) antibodies may also be used  
 CC as diagnostic agents for detecting the presence of TCAPs in samples  
 CC (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512  
 CC and AA24494 to AA24523 represent nucleotide and amino acid sequences  
 CC given in the exemplification of the present invention.  
 CC  
 XX  
 XX Sequence 480 BP; 78 A; 154 C; 141 G; 107 T; 0 other;  
 SQ

Query Match 44.28; Score 403; DB 22; Length 480;  
 Best Local Similarity 91.2%; Pred. No. 1,9e-112;  
 Matches 468; Conservative 0; Mismatches 0; Indels 45; Gaps 2;

OY 399 caagcagcaagcaactccctcgggtgagagatggagacttattcgttaagacatcggc 458  
 DB 480 CAGGCAAGCAAGCACTGCTCTCGGTGAGCATGGACCTTATTCGTAAGACATCAGGC 421  
 OY 459 tccagatagtaacttcagcagaagcgttcgcggagcaagaagagacagaagctaga 518  
 DB 420 TCCAGATATGAACTTTCAGCAGAGAGCGCTTGGCGGAGCAAGGGAGCAAGAAAGCTGAGA 361  
 OY 519 tgaacagctgcttgagcaaatcagcagcggaaggtgtctccgagctgcaccccg 578  
 DB 360 TGAACAGTGTGCTGCGACCAATTCACAGCGGGCAAGGGTGTCTCCGACCTCGCATCCCCCG 301  
 OY 579 gcgggggagcagctgaggtgctcccaagaagtgcattctgcttcgtcgaagggttgaaac 638  
 DB 300 GCGGGGGGCGAGCTGAGAGTCTCTCAGAGGTGCAATTCCTCTCCGAGGGGCTTGAAAC 241  
 OY 639 accaagagcactcagagatccttgagctcaagaagcagcagcccggttctcactccttg 698  
 DB 240 ACCAAGGACATCCAGAGGATCTCTGAGTCAAGCAGCAGCCCGGTTTCTTCACACTCCTTG 181

OY 699 gggtagacatggggtagccgcaatccaccctgtccttgctgctgcaagcagcactglttg 758  
 DB 180 GGGTAGACATGGGGGTAGCCGACAGCCCTGTCTTGCTGTGGACAGGCACTAGGTTTG 121  
 OY 759 cagctgtcccaagacaagccctgtcagctgcagagagccctgtcttgagagggccacgta 818  
 DB 120 CA-----GACAGGCCACAGTA 105  
 OY 819 ctctctcagcagagcttgagagcagaagcagcagaccagccacatgcaagcgtc 878  
 DB 104 C-TCTCAGCAGAGCTGTGAGAGCAGCAGGACAGACACCCGACGATGCAGAGGCTC 46  
 OY 879 tggcagcattgaccacccgtgtgctcggagcgc 911  
 DB 45 TGGCAGCATGACACACCGTGGGCTCCGGAGCGC 13

RESULT 5  
 AAF21864/c  
 ID AAF21864 standard; DNA; 494 BP.  
 XX  
 AC AAF21864;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Human breast and ovarian cancer associated antigen gene SEQ ID 251.  
 XX  
 KW Human; breast cancer; ovarian cancer; cytostatic; immunosuppressive;  
 KW neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 KW antidiabetic; antiinflammatory; antitumor; antileukemic; anticonvulsant;  
 KW antibacterial; antifungal; antiparasitic; cardiant; immune disorder;  
 KW Addison's disease; allergy; autoimmune hemolytic anaemia;  
 KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;  
 KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;  
 KW cardiovascular disorder; wound healing; neurological disease; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200055173-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000MO-US05881.  
 XX  
 PR 12-MAR-1999; 9905-0124270.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2000-611515/58.  
 DR P-PSDB; AAB58961.  
 PT  
 PT New human breast and ovarian cancer associated gene sequences and the  
 PT polypeptides encoded by these genes, useful in the prevention,  
 PT treatment and diagnosis of cancer, immune disorders, cardiovascular  
 PT disorders and neurological diseases -  
 XX  
 XX Claim 1; Page 673; 1299pp; English.  
 PS  
 XX Sequences AAF21614 - AAF22031 represent DNA sequences encoding human  
 CC proteins AAB58711 - AAB59128. The DNA and protein sequences are  
 CC associated with breast and ovarian cancer. Included in the invention are  
 CC sequences AAF22032 - AAF22040 and AAB59129 which are used in the  
 CC isolation and characterisation of the DNA and protein sequences of the  
 CC invention. The breast and ovarian cancer associated DNA, protein, agonist  
 CC or antagonist sequences exhibit cytostatic; immunosuppressive;  
 CC neurotropic; neuroprotective; antiviral; antiallergic; hepatotropic;  
 CC antidiabetic; antiinflammatory; antitumor; antileukemic; anticonvulsant;  
 CC antibacterial; antifungal; antiparasitic and cardiant activity. The  
 CC polynucleotide and protein sequences are used in the diagnosis of cancer,  
 CC particularly breast and ovarian cancer. The nucleic acid sequences,

CC proteins, agonists and agonists may also be used in the diagnosis,  
 CC prevention and treatment of immune disorders e.g. Addison's disease,  
 CC allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis,  
 CC diabetes mellitus, Crohn's disease, multiple sclerosis, Rheumatoid  
 CC arthritis and ulcerative colitis; cardiovascular disorders such as  
 CC myocardial ischemias; wound healing; neurological diseases such as  
 CC cerebral anoxia and epilepsy; and infectious diseases.

SO Sequence 494 BP; 121 A; 143 C; 132 G; 97 T; 1 other;

Query Match 41.6%; Score 378.8; DB 21; Length 494;  
 Best Local Similarity 90.1%; Pred. No. 4.5e-105;  
 Matches 445; Conservative 1; Mismatches 3; Indels 45; Gaps 2;

QY 413 tgcctcggggtgagcagtgagacatttctgttaagacatcaggtccagatagaact 472  
 DB 450 TTTTTCGGGTGGAGCAGTGGACCTTATTCGTTAAGACATCAGGCTCCAGATATGAAC 391  
 QY 473 ttccagcaagaagcgtctgcggggaagaaggaagcaagaagctgatatgaagctctg 532  
 DB 390 TTTCAGCACAAGCGCTTGGCCGGGAGCAAAAGGACAGAAAAGCTGAGATACAGTCTGG 331  
 QY 533 cagcaatcacagccggcggaaggtgtctcgagcctcgatcccccgccggcgggagctg 592  
 DB 330 CAGCAATCACAGCCGGCAGAGGCTCTCCGAGCCTCGCATCCCCCGCGGGGAGCTG 271  
 QY 593 gaggtgctcagaaggtgcatcttctgcttcctcgagggagcttgaacaccagaagctcca 652  
 DB 270 GAGGCGCTTCAGAAAGTGCATTCTCTTCCTGACAGGGGCTTGAAACACCAAGCACTCCA 211  
 QY 653 ggaatccctgagatcaaaagcaagaccgagctgtgtcactcctctgggggtgacatg 712  
 DB 210 GGGATCCTCGAGATCAAAACAGACAGCCCGGTTGTGCACTCTTGGGGTGACATGGGG 151  
 QY 713 tagccagagtcacacctgtcctgtgctgagcaagcagacatggttcagctgtccagac 772  
 DB 150 TAGCCGAGTCACACCTGTCTTGCTGGTGACAGGCACACTGGTTGCA----- 101  
 QY 773 aaagccctgtagctgcagagacccctgtctggaagagcccaagctactctctcagag 832  
 DB 102 -----GACAGGCCACAGTAC-TCTTCACGACAG 76  
 QY 833 ctggaagcagcaagagccagagaccagccagatgcagagcgtctgcaagccatgacc 892  
 DB 75 CTGGAGGAGCAGACAGGACGAGCAGCCCGCAGCATGACAGATGCTCTGGCAGCCATGACC 16  
 QY 893 accgtgggctccg 906  
 DB 15 ACCGTGGGCTCCGG 2

RESULT 6  
 AAT45887/c  
 ID AAT45887 standard; cDNA; 570 BP.

XX AAT45887;

XX 13-MAR-1997 (first entry)

XX Human colon specific gene CSG8 cDNA full-length clone.

XX Colon specific gene; CSG8; colon cancer; metastasis; diagnosis;

XX gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

XX CDS 125..370  
 XX /\*tag= a

XX W09639419-A1.

PD 12-DEC-1996.  
 XX 06-JUN-1995; 95WO-US07289.  
 XX 06-JUN-1995; 95WO-US07289.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX Rosen CA, Yu G;  
 XX  
 XX WPI: 1997-043054/04.  
 DR P-PSDB; AAM06550.  
 DR  
 XX  
 PT Human colon specific genes and their expression products - detection  
 PT of which, in non-colon tissue samples, can be used as indication of  
 PT colon cancer metastasis

XX Claim 15; Fig 8; 60pp; English.

CC 13 cDNA clones (AAT45880-92), only 2 of which are full length,  
 CC correspond to human colon specific genes, designated CSG1, CSG2,  
 CC etc., that are primarily expressed in tissues derived from the  
 CC colon. CSG7 and CSG10 show reduced expression in colon cancer  
 CC cells as compared to that in normal cells; the remaining genes are  
 CC overexpressed in colon cancer. The full-length sequences can be  
 CC used to isolate genomic clones including the complete gene. CSG  
 CC nucleic acids can be used to produce CSG polypeptides (see also  
 CC AAM06545-53) in transformed host cells, as probes to detect disorders  
 CC of the colon, partic. colon cancer and colon cancer metastasis, and  
 CC in gene therapy.

SO Sequence 570 BP; 126 A; 171 C; 160 G; 113 T; 0 other;

Query Match 33.5%; Score 305.4; DB 18; Length 570;  
 Best Local Similarity 85.7%; Pred. No. 1e-82;  
 Matches 431; Conservative 0; Mismatches 16; Indels 56; Gaps 6;

QY 413 tgcctcggggtgagcagtgagacatttctgttaagacatcaggtccagatagaact 472  
 DB 552 TTTTTCGGGTGGAGCAGTGGGAC-----TAGCTTAAACATCAGGCTCCAGATATGAAC 499  
 QY 473 ttccagcaagaagcgtctgcggggaagaaggaagcaagaagctgatatgaagctctg 532  
 DB 488 TTTCAGCACAAGCGCTT-CCGGAGCAAAAGGACAGAAAAGCTGAGATACAGTCTGG 440  
 QY 533 cagcaatcacagccggcggaaggtgtctcgagcctcgatccccgggcc-ggggcagct 591  
 DB 439 CAGCAATCACAGCCGGCAGAGGCTCTCCGAGCCTCGCATCCCCCGCGGGGAGCT 380  
 QY 592 ggaagtgctcagaaggtgcatctgctt---ccgcaggggttgaaccccaagcac 648  
 DB 379 GGAGGTGCTCTCAGAAAGTGCATTCTGCTTCCCTGTCAGGGGCTGAAACACCAAGGAC 320  
 QY 649 tccaaggaatcctgagatcaaaagcaagaccgagctgtgtcactcctctgggggtgacatg 708  
 DB 319 TTCAGGATCTCTGAGTCAAAACAGACAGCCCGGTTGTTCATCTCTTGGGGGTACATG 260  
 QY 709 ggggtagccgaagtcacacctgtcctgtgctgagcaagcagacatggttcagctgtccc 768  
 DB 259 GGGGTAGCCGCGAGTCCACCTCTGCTTGCTGGTGACAGGCACACTGTTGCA----- 206  
 QY 769 agacaagacctgtcagctgcagagaccctgtctggaagcagggccagctactctcagc 828  
 DB 207 -----GACAGGCCACAGTAC-TCTTCAGC 185  
 QY 829 agagctgagagcagcaagcagcagcagccagccagcagctgagagcgtctgagcagcat 888  
 DB 184 AGAGCTGAGAGCAGACAGCAGGCAAGCAGCAGCCCGCAGCATGACAGGCTCTGGCAGCAT 125  
 QY 889 gaccacgctgggctccggagcgc 911  
 DB 124 GACCACGCTGGGCTCCGGTACGC 102



Best Local Similarity 91.6%; Pred. No. 1.2e-70;  
Matches 318; Conservative 0; Mismatches 16; Indels 13; Gaps 3;

QY 418 tcgggtgagatgagacgttattcgttaagacatcagctccagatataacttcaag 477  
DB 399 TCGGCTGAGCATGGGACCTTATTGTTAAGACATCAGGCTCCAGATATGAATTTTCAG 340  
QY 478 cagaagcgttgcggcgagcaaaagagacagaagaactagatgaacagtgcctgacaga 537  
DB 339 CAGAA-ACCTTGCCCGGAGCAAAAGGACAGAAAACTGAGTGAACATGCTGCTGCACGA 281  
QY 538 atcaagccggcgcaagggtgtctcgaagctcgatcccccgcgcgggcgagctggaagt 597  
DB 280 ATCACAGCCGGGCAAGGCTGCTCAG-----CCTGCATCCCGAGGGGAGCTGAGAGT 230  
QY 598 gcccaagaagtgatctgtcttc---tgcagggtctgtaaaccaagcactccagg 654  
DB 229 GCCTCAGAAAGTGATGCTTCTCTAGTCAGGGGCTTGAACACCAAGGCATCCACAG 170  
QY 655 gatcctgagtcacaaagcagcagcccggtgtgtgacctccttggggtgacatgggta 714  
DB 169 GATCCTGAGTCAAAAGCAGCAGCCCCGGTTGTGCATCTCTTGGGGGTGACATGGGGTA 110  
QY 715 gccgagtcacacctgtcttctgtgctgacgagcagcactggtttgag 761  
DB 109 GCCGCAGTCCACCTGTCTTGGCCGCGACGACACATGTTTCAG 63

RESULT 9  
AAT8039/c  
ID AAT8039 standard; cDNA; 403 BP.  
AC AAT8039;  
XX  
XX 27-APR-1998 (first entry)  
DE Human intestinal trefoil factor (hTF) encoding cDNA.  
XX  
XX Trefoil factor; intestinal; hTF; human; inhiblt; lesion; ulcer;  
KW inflammation; cancer; treatment; ss.  
XX  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT CDS 2..226  
FT /\*tag= a  
FT /product= "intestinal trefoil factor"

PN MO9738712-A1.  
XX  
XX 23-OCT-1997.  
PD  
XX  
XX 11-APR-1997; 97WO-US06004.  
PF  
XX  
XX 12-APR-1996; 96US-0631469.  
PR  
XX  
XX (GEHO ) GEN HOSPITAL CORP.  
PA  
XX  
XX Podolsky DK;  
PI  
XX  
XX WPI; 1997-526205/48.  
DR P-PSDB; AAM27631.  
DR  
XX  
XX Trefoil polypeptide to treat or inhibit lesion formation - useful to  
PT treat or protect tissues against wounds, e.g. ulcers, inflammation,  
PT cancer and other insults  
XX  
XX  
XX  
PS Disclosure: Fig 6; 65pp: English.

XX This cDNA encodes a human intestinal trefoil factor (hTF). A clone  
CC comprising the DNA sequence of rat hTF is used to identify a cDNA clone  
CC encoding this hTF in a human intestinal cDNA library. The trefoil  
CC polypeptide, or a fragment of it can be used to treat or inhibit lesion

CC formation. The hTF polypeptide can be used to treat or protect against  
CC eye, skin or gastrointestinal or genitourinary tract wounds, e.g.  
CC lesions, ulcers, burns, inflammations or abrasions. It can also be used  
CC to treat non-ulcer dyspepsia, gastritis, peptic or duodenal ulcer,  
CC gastric cancer, MALT lymphoma, Menetrier's syndrome, gastro-oesophageal  
CC reflux disease, Crohn's disease, inflammatory bowel disease or neoplastic  
CC cancer. It can be used to protect tissues against injury caused by  
CC radiation, bacterial infection, alcohol or drugs. It can also be used  
CC for antibody production and diagnosis.

XX  
XX Sequence 403 BP; 74 A; 123 C; 111 G; 95 T; 0 other;

Query Match 29.1%; Score 265.4; DB 18; Length 403;  
Best Local Similarity 91.6%; Pred. No. 1.2e-70;  
Matches 318; Conservative 0; Mismatches 16; Indels 13; Gaps 3;

QY 418 tcgggtgagatgagacgttattcgttaagacatcagctccagatataacttcaag 477  
DB 399 TCGGCTGAGCATGGGACCTTATTGTTAAGACATCAGGCTCCAGATATGAATTTTCAG 340  
QY 478 cagaagcgttgcggcgagcaaaagagacagaagaactagatgaacagtgcctgacaga 537  
DB 339 CAGAA-ACCTTGCCCGGAGCAAAAGGACAGAAAACTGAGTGAACATGCTGCTGCACGA 281  
QY 538 atcaagccggcgcaagggtgtctcgaagctcgatcccccgcgcgggcgagctggaagt 597  
DB 280 ATCACAGCCGGGCAAGGCTGCTCAG-----CCTGCATCCCGAGGGGAGCTGAGAGT 230  
QY 598 gcccaagaagtgatctgtcttc---tgcagggtctgtaaaccaagcactccagg 654  
DB 229 GCCTCAGAAAGTGATGCTTCTCTAGTCAGGGGCTTGAACACCAAGGCATCCACAG 170  
QY 655 gatcctgagtcacaaagcagcagcccggtgtgtgacctccttggggtgacatgggta 714  
DB 169 GATCCTGAGTCAAAAGCAGCAGCCCCGGTTGTGCATCTCTTGGGGGTGACATGGGGTA 110  
QY 715 gccgagtcacacctgtcttctgtgctgacgagcagcactggtttgag 761  
DB 109 GCCGCAGTCCACCTGTCTTGGCCGCGACGACACATGTTTCAG 63

RESULT 10  
AAAS7156/c  
ID AAAS7156 standard; cDNA; 403 BP.  
XX  
XX AAAS7156;  
AC  
XX  
XX 16-OCT-2000 (first entry)  
DE Human intestinal trefoil factor cDNA.  
XX  
XX  
XX Human, intestinal trefoil factor; hTF; antilucer; antinflammatory;  
KW antimicrobial; cytostatic; gastrointestinal motility enhancer;  
KW peptic ulcer disease; inflammatory bowel disease; anticancer;  
KW gastrointestinal tract protection; bacterial infection; radiation injury;  
KW neoplastic cancer; ss.  
XX  
XX  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT CDS 2..226  
FT /\*tag= a  
FT /product= "hTF"

PN US6063755-A.  
XX  
XX 16-MAY-2000.  
PD  
XX  
XX 07-JUN-1995; 95US-0476705.  
PF  
XX  
XX 02-FEB-1994; 94US-0191352.  
PR 13-FEB-1992; 92US-0837192.

PR 25-MAR-1993: 9305-0037741.  
PR 14-FEB-1991: 9105-0655965.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
XX  
PI Podolsky DK;  
XX  
DR WPI: 2000-364484/31.  
DR P-PSDB; AAY99888.  
XX  
PT New rat intestinal trefoil factor, useful for treating peptic ulcer  
PT diseases, or inflammatory bowel diseases and for protecting the  
PT intestinal tract from injury caused by bacterial infection, or  
PT radiation injury -  
XX  
PS Disclosure; Fig 6; 17pp; English.  
XX  
CC The present sequence encodes human intestinal trefoil factor (hITF).  
CC It was isolated from a human colon cDNA library. The library was  
CC screened using a probe made from a fragment of human cDNA which had been  
CC amplified from human colon library cDNA using probes based on the  
CC sequence of rat intestinal trefoil factor. ITF may be used for the  
CC treatment of peptic ulcers and inflammatory bowel disease, and for  
CC protection of the intestinal tract from injury caused by bacterial  
CC infection or radiation injury. ITF may also be used to produce  
CC monoclonal antibodies for the detection of ITF in an intestinal tissue  
CC or blood serum by indirect immunosay. ITF can also be used to treat  
CC neoplastic cancer.  
XX  
SQ Sequence 403 BP; 74 A; 123 C; 111 G; 95 T; 0 other;

Query Match 29.1%; Score 265.4; DB 21; Length 403;  
Best Local Similarity 91.6%; Pred. No. 1.2e-70;  
Matches 318; Conservative 0; Mismatches 16; Indels 13; Gaps 3;

QY 418 tcgggtgagatgagacattatcgttaagacatgagctccagatgaaattcag 477  
DB 399 TCGGGTGGAGCATGGACCTTATTCGTTAAGACATGAGCTCCAGATATGAACTTCAG 340  
QY 478 cagaagagcttgcggagagaagaagagagaaagctgagatgaaagctgctgagca 537  
DB 339 CAGAA-AGCTTGGCCGGAGCAAAAGGACAGAAACTGAGATGAACGTGCTGGCAGCA 281  
QY 538 atcacagcgcggagaggtgtccgagcctcgatcccccgggggagcagctggaggt 597  
DB 280 ATCAACAGCCGGGCAAGGTGCTCAG-----CTGCAATCCAGGGGAGCTGGAGGT 230  
QY 598 gctcagaaggtgcatctgtctcc---tgcaagggtctgaacaccaaaggcactcag 654  
DB 229 GCGTCAGAAAGGTGATCTGCTCTCTAGTCAGGGGCTTGAACACCAAGCACTCCAG 170  
QY 655 gatcctgagatcaagaagcagcccggtgtgtgactcccttggagggtgacatgggta 714  
DB 169 GATCTTGAGTCAAGAGCAGCCCGGTGTGTCTCACTCTGGGGGTGACATGGGGTA 110  
QY 715 ggcgcacacacccctgtcctgtgctgagcagcagcactggttcag 761  
DB 109 GCGCGACTCCACCTGTCTTGTGGCGGACGCGACACTGTTTCAG 63

RESULT 11  
AAT22480/c  
ID AAT22480 standard; cDNA to mRNA; 263 BP.  
XX  
AC AAT22480;  
XX  
DT 22-AUG-1996 (first entry)  
XX  
DE Human gene signature HDMGS04091.  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;

KW cell typing; abnormal cell function; ss.  
OS Homo sapiens.  
XX  
XX WO9514772-A1.  
PN  
XX  
XX 01-JUN-1995.  
PD  
XX  
XX 11-NOV-1994; 94MO-JP01916.  
PF  
XX  
XX 12-NOV-1993; 93JP-0355504.  
PR  
PA (MATS/) MATSUBARA K.  
PA (OKUBA/) OKUBO K.  
XX  
XX Matsubara K, Okubo K;  
XX  
DR WPI: 1995-206931/27.  
XX  
XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
PS Claim 1; Page 1138; 2245pp; Japanese.  
XX  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T2637 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
SQ Sequence 263 BP; 47 A; 80 C; 70 G; 64 T; 2 other;

Query Match 28.5%; Score 259.4; DB 16; Length 263;  
Best Local Similarity 98.9%; Pred. No. 6.3e-69;  
Matches 260; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 396 tctcagcagaaagactgtcctcgtggtgagcatgagaccttattcgttaagacatca 455  
DB 263 TTTCAGGCAAGAAAGAACTGTCTCGGTGAGCATGGACCTTATTCGTTAAGACATCA 204  
QY 456 ggtcctagatatgaacttcaagcaagaagcttgcggagcaaaaggagcagaagaactg 515  
DB 203 GCGTCCGATATGAACCTTTCAGCAAGAGCCCTTCCGGGAGCAAAAGGCAAGAAAGCTG 144  
QY 516 agattgaacagtgtcctgccaacatcacagcgggcaaaagggtgtccgaagcctgcatccc 575  
DB 143 AGATGAAACAGTGTCTGACACAAATCACAGCCGGGCAAGGTTGCTCCAGGCTGCATCCC 84  
QY 576 ccggccggggcagctggaaggtgctcctcagaagggtgcatctgtcctcgaagggtctga 635  
DB 83 CCGGCCGGGGGCAAGCTGAGAGTGCTCTANAAAGMTGATTTGCTTCCTCAGGGGCTTGA 24  
QY 636 aacacaaaggcactcagagatc 658  
DB 23 AACACCAAGGCACTCCAGGATC 1

RESULT 12  
AAH34957/c  
ID AAH34957 standard; cDNA; 594 BP.

[illegible]

Db	354	TTGCACTCTCTTGGGGGTGACATGGGGGTAAGCGCAGTCCACCTGTCTTGGCTTGCGACG	295
Qy	746	gcacacttggcttgcagagctgtccacagacaagccctgtacgtgcagagccctgtctggg	805
Db	294	GCACACTGCTTTGCA-----G	279
Qy	806	acagagcccaactctactcctcaagacagctgtagagacaagaccagaccagcccccagc	865
Db	278	ACAGGCTCCACGATAC-TCTTCACGACAGCGTGGAGGACAGCAAGGCCACGACGCCACG	220
Qy	866	atgcagagcgcctcttgcagccatgaccaccgttggctccggagcgc	911
Db	219	ATGCAGAGCGCTGTGGCAGCCATGACCAACCGCGGGCTCCGGACGCG	174
RESULT 13			
AAV61325/G			
ID	AAV61325	standard; cDNA; 250 BP.	
XX			
AC	AAV61325;		
XX			
XX			
DT	.06-JAN-1999	(first entry)	
XX			
DE	Extended cDNA sequence of prostate tumour clone 1G-4736.		
XX			
XX	Prostate; cancer; tumour; vaccine; immunogen; clone; ss.		
KW			
XX	Homo sapiens.		
OS			
PN	W09837093-A2.		
XX			
PD	27-AUG-1998.		
XX			
PF	25-FEB-1998; 98WO-US03492.		
PR	09-FEB-1998; 98US-0020956.		
XX	25-FEB-1997; 97US-0806099.		
PR	01-AUG-1997; 97US-0904804.		
XX			
PA	(CORI-) CORIXA CORP.		
XX			
PI	Dillon DC, Xu J;		
XX			
DR	WPI: 1998-609886/51.		
XX			
PT	Polypeptides comprising immunogenic portions of prostate proteins -		
PT	used in a vaccine for the treatment of prostate cancer		
XX			
PS	Claim 12; Page 109-110; 130pp; English.		
XX			
CC	The present sequence is a DNA which encodes an immunogenic portion		
CC	of a prostate tumour protein. The encoded immunogen, or the DNA itself,		
CC	can be used as a vaccine for the treatment of prostate cancer. The DNA		
CC	was identified by analysis of a subtracted cDNA library obtained by		
CC	subtracting a prostate tumour cDNA expression library with a normal		
CC	tissue cDNA library.		
XX			
SQ	Sequence 250 BP; 55 A; 75 C; 63 G; 57 T; 0 other;		
Query Match 25.9%; Score 236.2; DB 19; Length 250;			
Best Local Similarity 98.8%; Pred No. 6.9e-62;			
Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;			
Qy	413	tgtctcgggttgagacatggaaccttactcgttaagacatcaggtccagatatgaact	472
Db	241	TTTTTCGGGTGAGCATGGGACCTTTATTGCTTAAGACATCAGGCTCCAGATATGAAGT	182
Qy	473	ttcagcagaagcgcttgcgggagcacaagggacagaaaagctgagatgaacagtgcctgg	532
Db	181	TTTCAGCAGAAAGCGCTTGCCGGAGCAAAAGGACAAAGAAAGCTGAGATGACAGTGCCTGG	122







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:39:03 ; Search time 5438.31 Seconds

(without alignments)  
1800.082 Million cell updates/sec

Title: US-09-867-034-1

Perfect score: 911  
Sequence: 1 ttttttttttgcctgttt.....cacggtggcctccggagcgc 911

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estom:\*  
5: em\_estopl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estcl:\*  
11: gb\_estc2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rtd:\*  
20: em\_gss\_vtl:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	568.2	62.4	600	AM513333	xc043f10.x
2	562	578	10	AM769057	h158b06.x
3	485.6	53.3	500	AM769425	h156b06.x
4	482.6	45.0	515	AM510389	x198f12.x
5	411.4	45.2	10	AM007096	w109b07.x
6	408	44.8	430	AA633399	np69n05.s
7	402	44.1	402	BE002129	799f01.x
8	401.2	44.0	476	AI281282	qk72d10.x
9	396	43.5	461	AI143630	qk74c01.x
10	392.2	43.1	463	AA315762	EST187535
11	386.8	42.5	476	AA552443	nk15d09.s
12	385.4	42.3	390	AA314975	EST186849

13	383	42.0	466	10	AM167728	AM167728	xc048c08.x
14	382.6	42.0	773	11	BG542020	BG542020	602572132
15	379.2	41.6	460	11	BI114257	BI114257	602862247
16	377.2	41.4	447	10	AI332337	AI332337	qp97h12.x
17	371.4	40.8	438	10	AI346752	AI346752	qp52g07.x
18	370	40.6	468	11	BG529874	BG529874	602558970
19	367.8	40.4	470	10	AA580138	AA580138	nh77a12.s
20	367.4	40.3	476	10	AA974853	AA974853	co30g10.s
21	366.8	40.3	457	10	AA716097	AA716097	z963c06.s
22	364.4	40.0	394	10	AI983974	AI983974	w020h03.x
23	362.8	39.8	402	10	AI973218	AI973218	wf53c01.x
24	362.8	39.8	420	10	AI748945	AI748945	at38h11.x
25	358	39.3	422	10	AI762067	AI762067	w153b12.x
26	356.4	39.1	420	10	AA507878	AA507878	nh75a09.s
27	348.4	38.2	443	10	BE645173	BE645173	7e64d06.x
28	346.2	38.0	404	11	W60304	W60304	z2d9d08.s1
29	345	37.9	405	10	AA552321	AA552321	nk06h03.s
30	344	37.8	353	10	AI350070	AI350070	qo28b12.x
31	340.2	37.3	412	10	AM517117	AM517117	xp91f10.x
32	335.4	36.8	454	10	AA953818	AA953818	co38e03.s
33	334.2	36.7	441	10	AA631397	AA631397	np86e07.s
34	333.4	36.6	404	10	AA614335	AA614335	np49c10.s
35	330	36.2	407	10	AA808607	AA808607	ce56b08.s
36	327.4	35.9	470	10	AA632754	AA632754	np85g07.s
37	324.2	35.6	482	11	W60395	W60395	z2d9d08.r1
38	323.8	35.5	343	10	AI832569	AI832569	at70b08.x
39	323.2	35.5	433	11	N74131	N74131	za75h01.s1
40	320.4	35.2	487	10	AI985964	AI985964	wf79d08.x
41	317.6	34.9	338	10	AA472805	AA472805	xq21c03.x
42	317.2	34.8	451	10	AA554791	AA554791	nk29c08.s
43	317	34.8	323	10	AM291863	AM291863	ui-H-B12-
44	317	34.8	352	10	AA951538	AA951538	EST363608
45	314.8	34.6	318	10	AI459674	AI459674	at84c01.x

#### ALIGNMENTS

RESULT 1  
AM513333  
LOCUS  
DEFINITION  
xc043f10.x1 NCI-CGAP UT1 Homo sapiens cDNA clone IMAGE:2706763 3'  
similar to gb:L15203 INTERSINAL TREPOIL FACTOR PRECURSOR (HUMAN);  
mRNA sequence.

ACCESSION  
AM513333  
VERSION  
AM513333.1 GI:7151411  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
1 (bases 1 to 600)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)

JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: [cgap@fmail.nih.gov](mailto:cgap@fmail.nih.gov)  
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/INL at:  
<http://image.llnl.gov/image/html/iresources.shtml>  
Seq primer: -40UP from Gibco  
High quality sequence stop: 423.  
Location/Qualifiers

#### FEATURES

source  
1..600  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2706763"

BASE COUNT	158 a	144 c	144 g	153 t	1 others
ORIGIN					
Query Match	62.4%	Score 568.2	DB 10	Length 600	
Best Local Similarity	96.8%	Pred. No. 9.7e-129			
Matches 579	Conservative 0	Mismatches 19	Indels 0	Gaps 0	
					/clone_11b-"NCI_CGAP_U13" /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors" /lab_host="DH10B" /note="Organ: uterus; Vector: PCMV-SpORF6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 1158-014"
OY	16	tgcttgctcataa	tgcttgctcataa	tgcttgctcataa	tgcttgctcataa
DB	3	ttttgtttcatt	ttttgtttcatt	ttttgtttcatt	ttttgtttcatt
OY	76	aacagtcagcag	aacagtcagcag	aacagtcagcag	aacagtcagcag
DB	63	acagtcagcag	acagtcagcag	acagtcagcag	acagtcagcag
OY	136	ctgtcttcac	ctgtcttcac	ctgtcttcac	ctgtcttcac
DB	123	ctgtctttc	ctgtctttc	ctgtctttc	ctgtctttc
OY	196	tgcttcccg	tgcttcccg	tgcttcccg	tgcttcccg
DB	183	tgcttcccg	tgcttcccg	tgcttcccg	tgcttcccg
OY	256	gtcttgatc	gtcttgatc	gtcttgatc	gtcttgatc
DB	243	gtcttgatc	gtcttgatc	gtcttgatc	gtcttgatc
OY	316	tgagggccg	tgagggccg	tgagggccg	tgagggccg
DB	303	tgagggccg	tgagggccg	tgagggccg	tgagggccg
OY	376	taaacacac	taaacacac	taaacacac	taaacacac
DB	363	taaacacac	taaacacac	taaacacac	taaacacac
OY	436	cttattcgt	cttattcgt	cttattcgt	cttattcgt
DB	423	cttattcgt	cttattcgt	cttattcgt	cttattcgt
OY	496	gcaaaagc	gcaaaagc	gcaaaagc	gcaaaagc
DB	483	gcaaaagc	gcaaaagc	gcaaaagc	gcaaaagc
OY	556	tgcttcgag	tgcttcgag	tgcttcgag	tgcttcgag
DB	543	tgcttcgag	tgcttcgag	tgcttcgag	tgcttcgag
RESULT 2					
LOCUS	AM769057	578 bp	mRNA	EST	04-MAY-2000
DEFINITION	h158b06.x1	NCI_CGAP_C014	Homo sapiens	CDNA clone	IMAGE:3005363
ACCESSION	AM769057				
VERSION	AM769057.1	GI:7701084			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	EMBL:U01010; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),				

JOURNAL COMMENT	Tumor Gene Index
Unpublished (1997)	
Contact: Robert Strausberg, Ph.D.	
Email: cgaibp-remail.nih.gov	
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmer-Buck, M.D., Ph.D.	
CDNA Library Preparation: Life Technologies, Inc.	
CDNA Library Arrayed by: Greg Lennon, Ph.D.	
DNA Sequencing by: Washington University Genome Sequencing Center	
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: info@image.lnl.gov	
Possible reversed clone: polyT not found	
Seq primer: -400P from Glbco	
High quality sequence stop: 409.	
Location/Qualifiers	
1. .578	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:3005363"	
/clone_id="NCI_CGAP_Col4"	
/tissue_type="moderately-differentiated adenocarcinoma"	
/lab_host="DH10B"	
/note="Organ: colon; Vector: pCMV-SpORf6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"	
BASE COUNT	153 a 142 c 142 g 141 t
ORIGIN	
Query Match	61.7%; Score 562; DB 10; Length 578;
Best Local Similarity	98.3%; Pred. No. 3.2e-127;
Matches 568; Conservative %	0; Mismatches 10; Indels 0; Gaps 0;
QY 10	tttgcgttggtcatcatatgtttacttgataaagaacaaacccgggaatgtacaag 69
Db 1	tttgccctgtttgctcattatatttactgtatcaaaagaaacaaacccagattagtcacag 60
QY 70	tattgaacagtaagcagagatgtgttctgtgaataaaggaccatttggaaagacatttatt 129
Db 61	tattgaacagtaagcagagatgtgttctgtgaataaaggaccatttggaaagacatttatt 120
QY 130	gacctgtgtcttcaaccaagaagacttgtgatttttgaanaacttcaactgaatglat 189
Db 121	ggcttgctgtgttccaccaaagaaactgtgatttttgaanaacttcaactgaatglat 180
QY 190	ttttcttgacttcccgagggaagcggaacttacaatgttttccaggcttccctgtgaagt 249
Db 181	tttttcttgacttcccgagggaagcggaacttacaatgttttccaggcttccctgtgaagt 240
QY 250	gtgcagactctgatttcaaatatccttgcatagtacatcgaactctttaggaagcttttcc 309
Db 241	gtgcagactctgatttcaaaaatattcttgcatagtacatcgaactctttaggaagcttttcc 300
QY 310	tgaccttgaggctcctgggcagaactctccctctgacacacctcccgcccttcccaagcag 369
Db 301	tgaccttgaggctcctgggcagaactctccctctgacacacctcccgcccttcccaagcag 360
QY 370	cagaaataaagcacaactctgaaagatctcaggcagaagaactgtctcgtgttggagca 429
Db 361	cagaaataaagcacaactctgaaagatctcaggcagaagaactgtctcgtgttggagca 420
QY 430	ttggagaccttattctgtaagaacatcagggtcccaataatgaacttcaagcagaagcgctt 489
Db 421	ttggagaccttattctgtaagaacatcagggtcccaataatgaacttcaagcagaagcgctt 480
QY 490	ccggagacaaaggagcagaaagcttgatgtaaacagtgccctggcagaatcaacgcccgg 549
Db 481	ccggagacaaaggagcagaaagcttgatgtaaacagtgccctggcagaatcaacgcccgg 540
QY 550	caagggtgctcgaagcctcgatcccccggccggggggc 587

Db 541 CAAGATGCTTGAAGCTTCGATCCGACCGCCAGGAGGC 578

RESULT 3  
LOCUS AW769425 500 bp mRNA EST 04-MAY-2000  
DEFINITION hl56b06.x1 NCI\_CGAP\_C014 Homo sapiens cDNA clone IMAGE:3005171 3' similar to gb:LI5203 INTERINAL TREFOIL FACTOR PRECURSOR (HUMAN); mRNA sequence.  
ACCESSION AW769425  
VERSION AW769425.1 GI:7701456  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 500)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov  
Seq primer: -400P from GIBCO  
High quality sequence stop: 377.

FEATURES  
Source  
1..500  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3005171"  
/clone\_lib="NCI CGAP Col4"  
/tissue\_type="moderately-differentiated adenocarcinoma"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT 133 a 117 c 116 g 134 t  
ORIGIN

Query Match 53.3%; Score 485.6; DB 10; Length 500;  
Best Local Similarity 98.2%; Pred. No. 1.6e-108;  
Matches 491; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 10 ttgagctgttgcataatgtttactgtacaaagaacaaaccaggaatagatcaag 69  
Db 1 TTTGGCTGTTTGTATATGTTTACTGTACAAAGAAACAAACCAGGAATATGACAG 60

QY 70 tatggaagaagtagcagaggtgtgttgaataaaggaccacttggagaacagtttatt 129  
Db 61 TATTCGAACAGTAGCAGAGGTGTTGTAATTAAGACCACTTGAAGACAGTTTATY 120

QY 130 ggcgtgctgtcttccaaagaagaactgtgtatttttgaanaactctactaactaatat 189  
Db 121 GCGTTGCTGTCTTCACCAAGAAAGACTGTGATTTTGAACAACTTCTACCTGAATGAT 180

QY 190 ttttctgcttcccgaggaagcggcacttaacagtgtcttccaggtcttccctgtgacgttg 249  
Db 181 TTTTCTGCTTCCCGAGGAAGCGGCACTTACAGTGTCTTCAAGCTTCTTGACGTGG 240

QY 250 gtgcagagctgagatcaaatatctcttcacatgacgtcagctcccttaaggagctttccc 309  
Db 241 GTGCCAGTGTGATTCAAAATATCCTTCGATGCACTGACGCTCTTAGGAGGATCTTTCC 300

QY 310 tgccttagagccttgagcagacttcccttgacacacctccgcgcctctccacagacgag 369  
Db 301 TGCCCTTAGGCGCTTGAGAGACTCTCCCTGACACCTCCCGCCTCTCCACAGACGAG 360

QY 370 cagaataaagcacaaactcagaaagltcagggcaggaagactgtctcgggtgagaca 429  
Db 361 CAGAAATTAAGCACAACTCAGAAAGTCTCAGGACGACGAAGAACTGCTCGGTGAGACA 420

QY 430 tggagccttattcgtttaagacatcaggtcccaagatagactttcaagaagagcttg 489  
Db 421 TGGGACCTTTATTCGTAGACATCAGCTTCATATGAACTTTTACAGAGGCGCTTG 480

QY 490 ccgagcagaagaagagacagaa 509  
Db 481 CCGGACGCAAGAGGACACAA 500

RESULT 4  
LOCUS AW510389 515 bp mRNA EST 03-MAR-2000  
DEFINITION x198f12.x1 NCI\_CGAP\_Ut1 Homo sapiens cDNA clone IMAGE:2682767 3' similar to gb:LI5203 INTERINAL TREFOIL FACTOR PRECURSOR (HUMAN); mRNA sequence.  
ACCESSION AW510389  
VERSION AW510389.1 GI:7148467  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 515)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
Tumor Gene Index  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/lresources.shtml  
Seq primer: -400P from GIBCO  
High quality sequence stop: 310.

FEATURES  
Source  
1..515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2682767"  
/clone\_lib="NCI CGAP Ut1"  
/tissue\_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"  
/lab\_host="DH10B"  
/note="Organ: uterus; Vector: pCMV-SPORT6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 139 a 119 c 116 g 141 t  
ORIGIN

Query Match 53.0%; Score 482.6; DB 10; Length 515;  
Best Local Similarity 98.0%; Pred. No. 8.8e-108;  
Matches 499; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

QY 16 tgtgttcataatgtttactgtacaaagaacaaaccgggaatagatcaagataga 75  
Db 4 TTTTGTTCATATATGTTTACTGTACAAAGAAACAAACCAGGATAGTACAGATTGA 63

QY 76 acagtagcagagtggtgttgaataaagaccacttggagaagactttattgtcttg 135

D	b6		ACAGTAGGAGAGTGTGTGAATAAAGAACCATTGGAAAGACATTTATTCGCCTTG       	123
OY	136		cctgtcctccacaagaagacttgyatttttgaaaccttcaccggaatgatatttttc       	195
D	124		CCTCCTCCACCAAGAACACTGTGATTTTTTGAACAACCTCACCGAAATGATTTTTTTC       	183
OY	196		tgcctcccggaggagcgacctaagtatgcccaagcttcctcgtacqcygggtgcc       	255
D	184		TGCTTTCCCAGAGGAAGCGGCACTTACATGTTTCCTAGCGTTTCTCTGTACAGCGGGTGCCA       	243
OY	256		gtctggattccaatatccctltycaclybcactygaagcttccttagggagtlcttcctgcct       	315
D	244		GCTCGATTTCAAANAATCTTGCATGACACTGAGCGTCTTAAGGAGTCTTTTTCCTGCCCT       	303
OY	316		tgaagcctcgggcagactctccccgtaaacctcccgccctctcccaagacgaagaaa       	375
D	304		TGAGGCTCGGGCAGACTCTCCCTCGACACCCCTCCGCCCCCTTCCACAGACGACGAGAAA       	363
OY	376		ttaaagcaaacctcagaagaftctcaggacgaagaagactctcgtggcggacatcgggac       	435
D	364		TAAAGCAACAACCTCAGAGAGTCTCATGACAGAGAACTCTCTCCGGGTGGACATGGGAC       	423
OY	436		cttatcttgttaagacalcaygcctccagatafgaatcttcagcaagaacycttcgcggga       	495
D	424		CTTATATTCGTTAAGACATCAGGCTCCAGATFATGAACTTTTAgCAGAGACGGCTTGCTGSA       	483
OY	496		gcaaaggacagcaaaagctlgatgaataca 524       	
D	484		GCCAA-GGACAGAAAAAGCTGAGATTMAAA 511       	
R	E	S		
L	A	M007096		
D	F	I	NITION	
V	e	wf09b07.x1 NCI_CGAP_Co3 Homo sapiens cDNA clone IMAGE:2506933 3'		
A	C	A007096	mRNA sequence.	
V	E	R	SION	
K	E	yW	A007096.1 GI:5855874 EST.	
S	O	r	human. Homo sapiens	
O	R	nism	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
R	E	fERENCE	NCI-CGAP http://www.ncbi.nlm.nih.gov/hcelgap.	
A	H	THORS	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
T	I	TLE	Unpublished (1997)	
J	O	URNAL	Contact: Robert Strausberg, Ph.D.	
C	O	MENT	Email: cgaps-remail.nih.gov	
			Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D., , Ph.D.	
			CDNA Library Preparation: M. Bento Soares, Ph.D.	
			CDNA Library Arraying: Greg Lennon, Ph.D.	
			DNA sequencing by: Washington University Genome Sequencing Center	
			clone distribution: NCI-CGAP clone distribution information can be	
			found through the I.M.A.G.E. Consortium/LNLB at:	
			www.bio.lnl.gov/bdip/image/image.html	
			Seq primer: -40Up from Gldco.	

```

FEATURES
SOURCE
location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2506933"
/clone_lib="NCI_CGAP_Cc3"
/sex="pooled"
/tissue_type="colon"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was

```

BASE COUNT 108 a 99 c 94 g 120 t  
ORIGIN

Query Match	45.28;	Score 411.4;	DB 10;	Length 421;	.
Best Local Similarity	98.68;	Pred. No. 2.4e-90;			
Matches 415; Conservative	0;	Mismatches 6;	Indels 0;	Gaps 0	

OY	11	tttgccgtttgttcataaagtttactgtatacaagaacaacccaggatagtacaagt	70
	1		
Db	1	TTTTTTTTTTGTTTCATATATGTTTACTGTACAAAGAAACGAAACTCAGAGATATGTCAAAGT	60
OY	71	atgtaacagtagcgaagatggtgtgtgaataaagaagccactltgaaagacagtttattgt	130
	1		
Db	61	ATTGAACAGTAGAGAGAGTGGTGTGAATTAAGGACACTTTGGAAACAGATTTTATTTG	120
OY	131	gcttgacgtcttaccagaagaagacttgtatltttgaaacttactactaaatgtatt	190
	1		
Db	121	GCTTGCCTGCTTCCACCAAGAAAGACTTGATTTTGGAAACTTCTTACCTGGAATGATTT	180
OY	191	ttttctgcttcccgaggaagcggcaacttacaagltgtccctaagcttltcctltgaacgttgg	250
	1		
Db	181	TTTTTCTGCTTTCCCGAGGAAGCGGCACTTAAGATGTTCTTAGGTTTCTGTGACGTGGG	240
OY	251	tgcacagtcgtgattcaataatatccttgcattgcactgcagctccttaaggagatccttccct	310
	1		
Db	241	TGCCAGAGCTGGAATTCAAAATATCTCTTGCAATGCACTGCACACTCTTTAAGGAGACTTTTTCCT	300
OY	311	ggcccttaggcctggggaagactctcccccgtgaacacctccgcgcctctccacgaagcagc	370
	1		
Db	301	GCCTTTAGGCGCTGGGCAAGCTCTCCCTGTGACACCCTCCCGGCTCTCTCCACGAGCGACG	360
OY	371	agaaataaagcacacaacctcagaaagtcctcaaggcaggaagaaactgtctcctgggttggagcat	430
	1		
Db	361	AGAAATTAACACACAACCTCAGAAAGTCTCAGGCAAGCAAGAACTGTCTCCGTGGGTGAGCAT	420
OY	431	g 431	
	1		
Db	421	G 421	

RESULT	6
AA633399	
LOCUS	
DEFINITION	AA633399 430 bp mRNA EST 28-OCT-1997 np69n05.s1 NCL_CGAP_BR Homo sapiens cDNA clone IMAGE:1131609 3'
ACCESSION	similar to gb:U15203 INTERIML TREFOIL FACTOR PRECURSOR (HUMAN);,, mRNA sequence.
VERSION	AA633399
KEYWORDS	AA633399.1 GI:2555259
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 430)
JOURNAL	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg Ph.D.

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: [www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

Insert Length: 563 Std Error: 0.00  
Seq primer: -40ml3 fwd. Ef from Amersham  
High quality sequence stop: 423.  
Location/Qualifiers

FEATURES  
SOURCE

1..430  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1131609"  
/clone\_1lb="NCI CGAP Br2"  
/sex="female, pooled"  
/tissue\_type="breast"  
/lab\_host="DH10B"  
/note="Vector: p773D-Pac (Pharmacia) with a modified  
polylinker; 1st strand cDNA was prepared from pooled bulk  
breast tumor tissue, and was then primed with a Not I -  
oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
RI adaptors (Pharmacia), digested with Not I and cloned  
into the Not I and Eco RI sites of the modified p773  
vector. This library is the normalized version of  
NCI CGAP Br1.1. Library was constructed by Bento Soares  
and M. Fatima Bonaldo."

BASE COUNT 94 a 125 c 136 g 75 t  
ORIGIN

Query Match 44.8%; Score 408; DB 10; Length 430;  
Best Local Similarity 99.5%; Pred. No. 1.6e-89;  
Matches 430; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 419 cgggtgagcatgggacattatcgtaagacalcagctccagatagaatttcagc 478  
|||||  
Db 1 CGGGGGAGCATGGGACCTTATTCGTTAAGACATCAGCTCCAGATATGAACCTTTCAGC 60  
479 agaagcgcttcgggagcaagagagacaaaagctgtagaagacagtgctcgagcaaa 538  
|||||  
Db 61 AGAAGCGCTTCCGGAGCAAAAGGAGCAAAAGCTGAGATGACAGTCCCGCAGCA 120  
539 tcacagccgggcaaggggtgctcgaagctcgatcccccggccgggggagctgaggtg 598  
|||||  
Db 121 TCACAGCGGGGAGAGGAGTGTCTCCAGGCTCCGATCCCGCGGC -GGGCGAGCTGAGGTG 179  
QY 599 cctcagaagtgatctgctctcctcaggggcttgaacacccaaggaactcaggagatc 658  
|||||  
Db 180 CCTCAGAGGTGATCTGTCTCTGCGAGGGGCTTGAACCAAGGCACTCCAGGATC 239  
659 ctgagatcaaaagcagaccccggtgtgtgactccttgagggtgtagcgtgagcg 718  
|||||  
Db 240 CTGAGTCAAAAGCAGCAGCCCGGTTGTGCACTCTTG666GTACATGGGGGTAGCG 299  
QY 719 cagtcacccctgtccttgctgctgcaagcagacactgttgcagctgtcccaagacaagcc 778  
|||||  
Db 300 CAGTCCACCTGTCTGTGGTGGGACGAGGACACCTGTTGACGCTGCCAGACAAAGCC 359  
QY 779 ctgtcagctccagagacccctgtgtgagcagagccaglaactcctcagcagagctgag 838  
|||||  
Db 360 CTGTGAGCTGGCAGAGCCCTTGTCTGGAGAGGCGCACGCTAC -TCTTCAGCAGAGCTGGAG 418  
QY 839 gacagcaagggc 850  
|||||  
Db 419 GACAGCAAGGCC 430

RESULT 7  
BF002129 402 bp mRNA EST 06-OCT-2000  
LOCUS 7g99f01.x1 NCI\_CGAP\_Col16 Homo sapiens cDNA clone IMAGE:3314617  
DEFINITION mRNA sequence.  
ACCESSION BF002129  
VERSION BF002129.1 GI:10702404  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgabs-remail.nih.gov  
Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLT, send email to:  
info@image.lnl.gov  
Seq primer: -400p from Glenco.

FEATURES  
SOURCE

1..402  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3314617"  
/clone\_1lb="NCI CGAP Col16"  
/tissue\_type="colon tumor, RER+"  
/lab\_host="DH10B"  
/note="Organ: colon; Vector: p773D-Pac (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI.  
Plasmid DNA from the normalized library NCI CGAP Col16 was  
prepared, and ss circles were made in vitro. Following NRP  
purification, this DNA was used as tracer in a subtractive  
hybridization reaction. The driver was PCR-amplified cDNAs  
from a pool of 5,000 clones made from the same library  
(cloneids: 1057416-1061255, and 1144584-114551).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 107 a 97 c 87 g 111 t  
ORIGIN

Query Match 44.1%; Score 402; DB 11; Length 402;  
Best Local Similarity 100.0%; Pred. No. 4.8e-88;  
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 ttgtcgtgtgttcataatgtttactgtacaaagaaacaaacccaggaatagataag 69  
|||||  
Db 1 TTTGCTGTGTTGTCATATGTTTACTGACAAAGAAACCAACCCAGGAATAGTACAG 60  
QY 70 tattgaacagtagagagagtggtgtgaaataaaggacacttggaaagacgtttact 129  
|||||  
Db 61 TATTGAAACAGTAGCAGAGGTGTGTGAATAAAGACCCACTTTCGAAGACAGTTTATT 120  
QY 130 ggcctgtcgtcttcacaaagaaagctgtgattttgaagactctacacgaatgfat 189  
|||||  
Db 121 GCGTTGCTGTCTTCCACAAAGAAAGACTTGTGATTTTAAANACTTCTACCTGAATGTAT 180  
QY 190 ttcttcgtcttcacgaagagcagcacttacagtgctcctcagagctcttcgtgagctgg 249  
|||||  
Db 181 TTTTCTGCTTTCCCGAGGAAGCGGCACTTACAGTGTCTCAGAGCTTCCGTGACGTGG 240  
QY 250 gtgcagctctgattcaaatatcctgtcagcagcagctcagctcctctagagagctttcc 309  
|||||  
Db 241 GTGCCAGTCTGATTCAAATAATCTTCACATGACAGCTCCTTATGAGGAGTCTTTTCC 300  
QY 310 tgccttgaagccttggaagactctcccttgaacacccctccgacctcccaagagagcgag 369  
|||||  
Db 301 TGCCCTTGAAGGCTGTGGGAGACTCTCCCTGACACCCCTCCGCTCTCCACAGAGCGAG 360  
QY 370 cagaataaagcacacacactcagaaagtlctcagggcagagaaga 411  
|||||  
Db 361 CAGAAATAAAGACACACTCTCAGAAAGTCTCAGGACGAGAGAA 402

RESULT 8  
A1281282

LOCUS	AI281282	476 bp	MRNA	EST	23-NOV-1998
DEFINITION	ql12d10.x1 NCI CGAP Co8 Homo sapiens cDNA clone IMAGE:1874515 3' similar to gb:U5203 INTERIMINAL TREPOIL FACTOR PRECURSOR (HUMAN);, mRNA sequence.				
ACCESSION	AI281282				
VERSION	AI281282.1 GI:3919515				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 476)				
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .				
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index				
JOURNAL	Unpublished (1997)				
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://www-bio.llnl.gov/bbrp/image/image.html">www-bio.llnl.gov/bbrp/image/image.html</a> Seq primer: -400P from Glibco High quality sequence stop: 423. Location/Qualifiers 1. .476 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1874515" /clone_11b="NCI CGAP Co8" /tissue_type="adenocarcinoma" /lab_host="DH10B" /note="Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - ol19o(dt) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Donaldo."				
FEATURES	source				
BASE COUNT	106 a 138 c 148 g 84 t				
ORIGIN					
Query Match	44.0%; Score 401.2; DB 10; Length 476;				
Best Local Similarity	90.7%; Pred. No. 7.5e-89;				
Matches 468; Conservative	0; Mismatches 3; Indels 45; Gaps 2;				
QY	396 tctcagcagcagaagactgctcctcggtggtgagacatgagacattatctgtaagacatca 455				
Db	6 TTTTCAGGCACGAAAAAAGCTGCTCTCGGGGAGACATGGGACCTTTATTCGTTAAGACATCA 65				
QY	456 ggcctcagatatgaacttcagcagaagcgcgtccggaggaacaaaggagcagaaaagct 515				
Db	66 GGCCTCAGATATGAACTTTCAGCAGAAAGCGCTTCCGGGAGGCAAAAGGACAGAAAAAGCTG 125				
QY	516 agatgaacagatgacctggagagcaatcaacagccggaggaagggatgctcagagcctcgatcc 575				
Db	126 AGATGAACAGATGCCCTGGCAGCAATACAGCCGGGCAAGGGGTCTCCGAGACCTCGATGCC 185				
QY	576 ccgagccgggggcagcttggaagtgtcctcagaaggtgcatctcgtccctcgcagaaggtctga 635				
Db	186 CCGGCGGGGGGAGCTGGAGAGTCCCTCAGAAAGTGCATTTCCTTCCTCGACAGGGCTTGA 245				
QY	636 aacaccaaagcactcagggatccttggatcaaaacagcagagcccggtgtgtgactcct 695				
Db	246 AACACCAAGGCACTCCAGGAGTCTCGAGATCAAAACACAGACGCCGGTGTTCACACTCCT 305				
QY	696 tggggggtgacatgggggtgagccgagtcacacccgtctccttgctgagcaagcaactggt 755				

	Db	306	tggcggtgacatdggsgtgagccgcacttccacccttgccttggcgtggcacggcacacttgt	365		
Oy	756	ttgcagctgtgccagacaagaaccctgtlcaagtgtccagaagcccttgccttgtagaacaggcccc	815			
Db	366	TTCGA-----	-GACAGCCACC	381		
Oy	816	gtaccttcctcgaagagctgtgagagacagaaagcccagaccacccccagatatgcagagc	875			
Db	382	GTAAC-TCTCAGAGAGACTGGAGAACAGCAAGGCACAGGACCGCCCATGTCAAAGCG	440			
Oy	876	cctcgcagcatgaccaccgttggtgtccggagcgc	911			
Db	441	CCTCGCACCATGACCCACCGTGCGTCCGCCGACGC	476			
RESULT	9	AII143630	461 bp	mRNA	EST	26-OCT-1998
LOCUS	AII143630	gbJ4c01.x1 Soares_fetal_heart_NbHn19w Homo sapiens cDNA clone				
DEFINITION	IMAGE:I705824.3,	similar to gbJL15203	INTestinal TREPOIL FACTOR PRECURSOR (HUMAN);			
ACCESSION	AII143630					
VERSION	AII143630.1	GI:3665439				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
TITLE	1 (bases 1 to 461) NCI-GAP http://www.ncbi.nlm.nih.gov/ncigap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Human Gene Index Unpublished (1997)					
JOURNAL	Contact: Robert Strausberg, Ph.D. Email: cgaphs-r@mail.nih.gov This clone is available royalty-free through LNL; contact the IMAGE Consortium ( <a href="http://info@image.lnl.gov">info@image.lnl.gov</a> ) for further information. Insert Length: 510 Std Error: 0.00 Seq primer: -40m13 fwd. Et from Amersham High quality sequence stop: 416. Location/Qualifiers 1. 461					
FEATURES	/organism="Homo sapiens"					
source	/db_xref="taxon:9606"					
	/clone="IMAGE:I705824"					
	/clone_lib="Soares_fetal_heart_NbHn19w"					
	/sex="unknown"					
	/dev_stage="19 weeks"					
	/lab_host="DH10B (ampicillin resistant)"					
	/note="Organ: heart; Vector: pRTD3 (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer (5 TGTATGCATCTGAGATGGAGCCGCCGCATGTTTTTTTTTTT 3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pRT vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Felipe Bonaldi. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19w."					
BASE COUNT	103 a	136 C	146 g	76 t		
ORIGIN						
Query Match	43.5%	Score 396;	DB 10;	Length 461;		
Best Local Similarity	91.1%;	Pred. No. 1.4e-86;				
Matches	461; Conservative	0; Mismatches	0; Indels	45; Gaps	2	
Oy	399	caaagcagaagacgtctcgtggtgagacatcgagacattatcgtttaagacataagg	458			
Db	1	CAGGACGACGAACTGTCTCTCGGCTGGACATAAGGACCTTTATTCTTTAAGACATAGGC	60			



QY	459	tcacgaatgaacttcacgaagaagccttgcgggagcaaaagagacaagaagcctgaga	518
Db	61	TCcgaATATGAcACTTTCAGcAGAAAGccCTTgcGGAGCAAAAGGAGcAGAAAAGCTGAGA	120
QY	519	tgaacagtgccttgcagcaatcacaagccggycaaagggtgtctccgagctctgcataccccc	578
Db	121	TGAACAGTGCCTGGcCAGCAATATCAAGCCGGcCAAGGGTGCTCCGAGCTCGATCCCCC	180
QY	579	gcccggggcgaacttgaggtgtgcctcaagaagtgcattctgtctctgaagggtcttgaac	638
Db	181	GCCGGGGcCACTGAGGTGCTTcAGAAAGTGcATTTGCTCTCTGcAGGGGCTTGAAC	240
QY	639	accaaggaaccacccagggatcccttgagtccaagcagacagccccgggtgttgtacatccttg	698
Db	241	ACCAAGGcACTCCAGGGATCTCTGcAGTCAAAAGcAGCAAGCCCCGGTTTGTACTCTTGG	300
QY	699	gggttgacatggggtaagccgagctccaccctgtctcttgctgycagagcaacatggttg	758
Db	301	GGGTGAcATGGGGGTAGCCCGAGTCCAACCCTGTCTTGGCTGGcACCGCACTGGATTG	360
QY	759	cagctgtccacagacaagcccttgcagctgcccagagcccttgcctgggaacagggcccaatga	818
Db	361	CA-----GACAGGCCCAAGTA	376
QY	819	cttcctcaagcagaactctgaaagagacagaagggccagagacacccacagatcaagaagcttc	878
Db	377	C-TCTCTAGcAGACTCTGGAGAGcAGCAAGGCCAGAGCACCCcAGCATCTAGAGCGCTC	435
QY	879	tgagcagcatatcacaccgcttggtcccc	904
Db	436	TGGcAGCCATGAcCAACCGTGGGCTCC	461

RESULT	10				
AA315762/c					
LOCUS	AA315762	463 bp	mRNA	EST	19-APR-1997
DEFINITION	EST187535 Colon carcinoma (HCC) cell line II Homo sapiens				CDNA 5
sequence.	end similar to similar to trefol factor, intestinal, mRNA				

ACCESSION	AA315762
VERSION	AA315762.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE
Homo sapiens	Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi	1 (bases 1 to 463)
Mammalia: Eutheria: Primates: Catarrhini: Homiidae: Homo.	

C.J., Lee, N.H., Kirkness, E.F., Weinstein, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Whai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitzhugh, W.M., Frichman, J.L., Geoghegan, N.S., Glodet, A., Grehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Behanik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.-F., Ferrie, A., Fisher, C., Hastings, G.A., He, M.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.

**TITLE**  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence

**JOURNAL**  
Nature 377 (6547 Suppl.), 3-174 (1995)

**MEDLINE**  
96026280  
Other-ESTs: THC178105

**COMMENT**  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlavetigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hg1/hg1.html>)  
Seq primer: M13 Reverse.  
Location/Qualifiers

```

SOURCE
1. .463
/organism="Homo sapiens"
/db_xref="ATCC (lnhost):111494"
/db_xref="taxon:9606"
/clone_id="Colon carcinoma (HCC) cell line II"
/tissue_type="colon"
/cell_type="KM12C"
/cell_line="KM12C(HCC)-parental human colon carcinoma
;Dukes B2"
/note="Organ: colon; Vector: pbluescript SK-; Site_1
ECORI: Site_2: XhoI"
BASE COUNT
77 a 146 c 134 g 105 t 1 others
ORIGIN

```

Query Match	43.1%;	Score 392.2;	DB 10;	Length 463;
Best Local Similarity	-90.4%;	Pred. No. 1.2e-85;		
Matches 459; Conservative	0;	Mismatches 4;	Indels 45;	Gaps 2;

OY	399	caacacgaagaacatgctctcgtgggttgagcaatggacattatctgtaaaacatcaagc	451
Db	463	CAGGACCAAGAATATGTCTCTGGGGTGAAGCATGGGACCTTTATTCGTTAAACATCAGGC	400
OY	459	tccagatatgaacttccagcaaaagcgtctgcgaggacaaaggacagaaagtga	511
Db	403	TCCAGATATGGAACCTTTTCAGCAGAAAGCGGCTTCGCGGAGCAAAAGGAGACGAAGAAAGCTGAGA	344
OY	519	tgaacatgaccttggcagaagaatcacagccgggcaaggagtgtgcctcgagacctgcatccccg	578
Db	343	TGAAACATGCTCTGGCAGCAATTCACAGCCGGGCAAGGGTGTCTCCGAGCTCCCATTCCCCCG	281
OY	579	gcgcggggcaactgtagagttgctctcaagaaagtgcattctgtctcttgcaggagcttgaac	631
Db	283	GCCGGGGGCACATCGAGGTGGCTCTCAGAAAGTGTGCAATTGCTCTCTGCGAGGGGCTTGAAC	224
OY	639	accaaaggaacctcagaaggatctctgagttcaaaagcagccccggttgttgaacctcttg	691
Db	223	ACCAAGGACATCCAGGGATCTCTGGAGTCAAAAGCAGACCCCCGGTATGTTGCACTCTTTGG	161
OY	699	ggatgacatatggggatagccgagttccacacctgtctcttgctgcagaggaacacactgattg	751
Db	163	GGGTGACATGGGGGAGGCCGAGTCCACCTGTCCCTTGGTGGCACGGCACATGTTGTTG	104
OY	759	cagctgttcccaagacaagacctgtgcagctgtgcagagacctgtctcgggacaaagccccagta	811
Db	103	CA-----GACAGGCCACAGTA	88
OY	819	ctctctcagcagaagctgtagagacacagaagcagagaccacagcatctgacagcctc	871
Db	87	C-TCTCTAGCAGAGCTGGAGAGACAGCAAGGCCAGAGACCACGCCAGCATGCAAGANGTC	29
OY	879	tggcagcgcattgacacacggttggtctccgg	906
Db	28	TGGCAGCCATGACCAACCGTGGGCTCCGG	1

RESULT	11		
AA552443			
LOCUS	AA552443	476 bp	mrna
DEFINITION	nm15009.s1 NCI-CGP_Co2 Homo sapiens CDNA clone IMAGE:101355 3' similar to gb:115203 INTESTINAL TGFBI-F1 FACTOR PRECURSOR (HUMAN); mRNA sequence.	EST	05-SEP-1997
ACCESSION	AA552443		
VERSION	AA552443.1	GI:2322697	

KEYWORDS	Est.	human.
SOURCE	ORGANISM	Homo sapiens
REFERENCE	1 (bases 1 to 476)	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	NCI-CCAP	http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs@email.nih.gov Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D. CGMA Library Preparation: Stratagene, Inc., David B. Krizman, Ph.D. CGMA Library Arraying: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILMIL at: www-bio.llnl.gov/bbrp/image/image.html Insert Length: 666 Std Error: 0.00 Seq primer: ~40m3 fwd. ET from Amersham High quality sequence stop: 321. Location/Qualifiers	
FEATURES	source	1..476
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:1013585"	
	/clone_id="NCI_CGAP_Co2"	
	/tissue_type="tumor"	
	/lab_host="SOLR (Kanamycin resistant)"	
	/note="Organ: colon; Vector: Bluescript SK-; Site.1: EcoRI ; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Bulk colon villous adenoma. 5' adaptor sequence: 5' GAATTCGGCAGCAG 3' 3' adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' Average insert size: 1.1 kb."	
BASE COUNT	107 a 136 c 152 g 81 t	
ORIGIN		
Query Match	42.5%, Score 386.8; DB 10; Length 476;	
Best Local Similarity	87.8%; Pred. No. 2.5e-84;	
Matches 453; Conservative	0; Mismatches 17; Indels 46; Gaps 1;	
QY	366 tctatagcagcaagaactgctctctggtggtgagcattggaactttatttgtaagacatca	455
Db	1 TTTTACGGCACAAGAACTGCTCTCGGGTGGAGCAAGGAGCACTTTATTTGTTAAGACATCA	60
QY	456 ggtccagatatagaacttccagcaagaacgcgttgcggagagcaaaagagcagaagctg	515
Db	61 GGGTCCACATATGAACTTTCAGCAGAAACCGCTGCGGGAGCAAAAGGACAGAAAGCTG	120
QY	516 agatgaacagtgcctggtgcagcaaatcaacgcccgggcaagggtgtctcgaagcctgcattccc	575
Db	121 AGATGAACAGTGCCTCGGCGAGCAATACACGCCGGGCAAGGAGTCTCCGAGCCTCGCATCCC	180
QY	576 ccggccggggggcagctgtgaagtgctccaaagaagtgatcatttcgtcttcgcagagggtctga	635
Db	181 CCGGCGGGGGCAGGCTGGAGGGTGCTTCAGAAAGTGTCATTCTGCTTCGCAAGGGGCTTGA	240
QY	636 aacacaaagcactcacaaggatcctcgtgagttcaaaacagacagccccgggttgtgcactct	695
Db	241 AACACCAAGGCACTCCAGAGATCCCGAGGTCAAAACACACAGCCCGGTTGTTGCACCTCT	300
QY	696 tgggggtgacatggtgggttagccgacagttccacccgtctcctgtgctgtgcaaggcaactggt	755
Db	301 TGGGGGTGACATGGGGGTAGCGCGCAGTCCACACCTTGTTGGCTGGCAGCGCACACTGTT	360
QY	756 ttgagcgtgtcccaagacaagaccctgttaagctgtccagagccctgtctgtgagacaggtccac	815
Db	361 TTTCAGACAGCGCCACGTA-----	378

OY	816	gtcttccatcagacgagtcttgaggacaacgaacgcccagaccagccccagcatctgcagacg	875
Db	379	-----CTCTCAGACGAGCTGGAGAGCACACCAGCCAGGACCAAGCATGCAGACCG	434
OY	876	ctctgcacgatcatgaccacgcgtgggtcttcgcggagacgc	911
Db	435	CTCTGGAAGCATATGACATCGTGGGGCTTCGGGAGCGC	470
RESULT	12		
LOCUS	AA314975/c		
DEFINITION	AA114975	390 bp mRNA	EST
VERSION	AA314975.1	GI:1967374	19-APR-1997
KEYWORDS	AA314975		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 390)		
AUTHORS	Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S., Glodak,A., Grehm,C., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmirois,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudke,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferlie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Ruben,S.M., Dillon,P.J., Fannon M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.		
TITLE	Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence		
JOURNAL MEDLINE	Nature 377 (6547 Suppl), 3-174 (1995)		
COMMENT	Other ESTs: THCI78105 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@vetvir.org For clone availability, additional sequence and expression information related to this EST, please check The TIGR Human Gene Index ( <a href="http://www.tigr.org/tdb/hgl/hgl.html">http://www.tigr.org/tdb/hgl/hgl.html</a> ) Seq primer: M13 Reverse. Location/Qualifiers 1..390 /organism="Homo sapiens" /db_xref="ATCC (inhost):113179" /db_xref="taxon:9606" /clone_lib="HCC cell line (metastasis to liver in mouse) II" /tissue_type="colon" /cell_type="KM12SM" /cell_line="KM12C(HCC)metastasis into mouse (liver)" /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"		
BASE COUNT	71 a	108 c	113 g 95 t 3 others
ORIGIN			





Sat Mar 30 14:59:28 2002

us-09-867-034-1.rst

Page 11

Search completed: March 30, 2002, 09:31:07  
Job time: 13924 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 09:32:17 ; Search time 5438.31 Seconds

(without alignments)  
5157.205 Million cell updates/sec

Title: US-09-867-034-5

Perfect score: 2610  
Sequence: 1 gatgtgcgcacgtccacag.....ccgagatggtcgcagagagt 2610

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST: \*  
1: em\_estfun: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estom: \*  
5: em\_estpl: \*  
6: em\_estda: \*  
7: em\_estro: \*  
8: em\_estov: \*  
9: em\_hic: \*  
10: qd\_estl: \*  
11: qd\_est2: \*  
12: qd\_hic: \*  
13: qd\_gss: \*  
14: em\_gss\_fun: \*  
15: em\_gss\_hum: \*  
16: em\_gss\_hiv: \*  
17: em\_gss\_pln: \*  
18: em\_gss\_pro: \*  
19: em\_gss\_rtd: \*  
20: em\_gss\_vit: \*  
21: em\_gss\_other: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	653.6	25.0	877	11	BE872086 601448177
2	555.4	21.3	952	11	BC386415 602455641
3	433.7	16.7	621	10	AW969591 EST381668
4	433.6	16.6	666	10	AI732648 nm73f01.x
5	368	14.1	536	10	AI801356 t088c05.x
6	358.6	13.7	525	10	AI222799 qp39d10.x
7	351	13.4	450	10	AW820993 RC2-ST030
8	343.4	13.2	1779	12	AK007445 Mus muscu
9	325	12.5	429	10	AI821412 nes8d09.x
10	320.6	12.3	517	10	AW610230 RC2-ST030
11	313.4	12.0	449	11	BF431333 naa43f01
12	312	12.0	312	10	AA328866 EST32671

C 13	308.2	11.8	418	10	AI281212	AI281212 GK58609.x
14	306	11.7	491	10	AI791944	AI791944 nm73f01.y
15	305	11.7	450	10	AA527898	AA527898 nes8d09.s
C 16	301	11.5	546	10	AW140090	AW140090 UI-R-Btl-1
C 17	280.6	10.8	470	10	AA576913	AA576913 nm73f01.s
C 18	256.4	9.8	374	11	W87535	W87535 zh55g04.s1
C 19	248.4	9.5	427	10	AI394677	AI394677 t924p01.x
C 20	225.6	8.6	443	11	BF000690	BF000690 7h34p01.x
C 21	192.8	7.4	226	10	AI739162	AI739162 w127a01.x
C 22	183.8	7.0	236	10	AA583007	AA583007 nm76f03.s
C 23	170.4	6.5	2237	12	AK011082	AK011082 Mus muscu
C 24	167.4	6.4	199	10	AW085900	AW085900 xc60f12.x
C 25	149.2	5.7	257	11	BF766225	BF766225 IL2-CS005
C 26	149.2	5.7	257	11	BF766229	BF766229 IL2-CS005
C 27	148.4	5.7	644	11	BF579351	BF579351 602095640
C 28	124.6	4.8	441	11	H80869	H80869 yu58h04.s1
C 29	115	4.4	707	10	AW390130	AW390130 CM4-ST017
C 30	108.2	4.1	515	10	BE232772	BE232772 137905 MA
C 31	106.8	4.1	191	10	AW376331	AW376331 RC6-CM020
C 32	92.8	3.6	569	10	BE235607	BE235607 143157 MA
C 33	88.6	3.4	960	11	BE788920	BE788920 601585470
C 34	86.8	3.3	883	11	BI257633	BI257633 602967550
C 35	83	3.2	822	11	BE896215	BE896215 601438928
C 36	81.4	3.1	631	10	BE280813	BE280813 601155253
C 37	81.4	3.1	682	11	BF196339	BF196339 7n68f01.x
C 38	81.4	3.1	873	11	BF984445	BF984445 602307990
C 39	79.8	3.1	744	11	BG327067	BG327067 602426275
C 40	79.6	3.0	941	11	BG341123	BG341123 602462772
C 41	78.8	3.0	561	11	BG229942	BG229942 mac24f09.
C 42	78.8	3.0	665	13	AZ305382	AZ305382 IM0006J04
C 43	78.2	3.0	504	11	BE901092	BE901092 601674449
C 44	77.2	3.0	376	10	AI578443	AI578443 UI-R-Y0-U
C 45	77.2	3.0	400	10	AI577576	AI577576 UI-R-Y0-U

#### ALIGNMENTS

RESULT 1  
BE872086 877 bp mRNA  
LOCUS 601448177F1 NIH\_MGC\_65 Homo sapiens CDNA clone IMAGE:3852018 5',  
DEFINITION mRNA sequence.  
ACCESSION BE872086  
VERSION BE872086.1 GI:10320862  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 877)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: gcaps-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: L1AM9573 row: 1 column: 19  
High quality sequence stop: 645.  
Location/Qualifiers  
1. 877  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3852018"  
/clone\_id="NIH\_MGC\_65"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"

#### FEATURES

source

```

/note="Organ: colon; Vector: pcMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT      180 a      261 c      300 g      136 t
ORIGIN

```

Query Match	25.0%;	Score 653.6;	DB 11;	Length 877;
Best Local Similarity	95.1%;	Pred. No. 9.3e-126;		
Matches 761;	Conservative 0;	Mismatches 29;	Indels 10;	Gaps 8;

LOCUS	DEFINITION	602455641f1	NIH_MCC_15	Homo sapiens	cdna	clone IMAGE:4583723	5'
QY	1190	ggaagaaagatgtctacagacgcacaaagtgatgctcttcaacagatgtcacaagatcaagccaag	1249				
Db	33	GGAGGAGAAAGTCTACGACGACCAAGTAGCTTCTACAAAGATGTCCAGATCAA-CCAGA	91				
QY	1250	gcaggtctgagatgaagaactctgtgagctctgagcgcaattcttcctcctgactgttccaag	1309				
Db	92	GCAGGTGTGAAGATGAACACTTTTGTGAGTCTGTAGGCCATTTCT-CTGACCTTGTCAAG	150				
QY	1310	aagatccgtgatttccacagaaataactcaacctccacccacagatgatagaatgattctca	1369				
Db	151	AAGATCCGTGATTCTTCACAGAAATAACTCACCTCCACAGATGATGAGATGTTCTC-	209				
QY	1370	agaaacttggtgcatactatcttgaaataagattcaggtgtcatcaactctgagacctcaag	1429				
Db	210	AGAAACTTGTCGCGCATCATATGTGAAATTAATTCAGGGGTCATCACTGTGGACCTTCAG	269				
QY	1430	cgcacagccgagaaacctgtgattctctcgaaagacagagaatcagtgagtgatacccgacga	1489				
Db	270	CGCCAGCCGGAG-CTGTGTTCTCTCGGAAGACAGGAAGTCATGAGTACACCCGGCAGA	328				
QY	1490	agaagaacctgtcaagacagacccctctgcctcgaacgacctcccgcggttcttggcttcc	1549				
Db	329	AGAAAGACCTTCACAGACAGCCCCCTCGCTTGACCGCTCCCGCGGTTCTGGGCTTCC	388				
QY	1550	cggagcttctctcccgggcgccacacgctcttgcaagtgtgaccttgagctcttggcgaaggcg	1609				
Db	389	CGGCTTCTCTCTCCGGGCGCCACCGCTGGAGGTTGACTGTCACTGGGCGACCGCGCG	448				
QY	1610	gtctcaagctgtgggtgtgcccggggaagggtgtgagagagacagagagatgtgaactcagc	1669				
Db	449	GCTGCACGATGTGGGGTGGCCGGGAGGGGGTGAAGAGGACAGGAGATGGACTCAGC	508				
QY	1670	gcgaggaagcggtctctggtgcgtgtgatactctctgtcaaccaagcagttgctggccagacct	1729				
Db	509	GCCAGAGACGGCGTGTGGGCGCGTGAATCAATCTC--GCACACAGCAGTGTGGGCGACGACT	566				
QY	1730	ccccgggcacagacactgtctgcgtgtgaacgagatcccgcggaagcgtgtgaagtgcacctga	1789				
Db	567	CCCCGGGCACCGACCTGT-CCGCTGTGACCGAGATCCCGCG-GGCTGTGAAGATCACTTGA	624				
QY	1790	ctacgaagcggggaagtgagctcctccacaacagcccagaagcccgaagggtccatctcctaac	1849				
Db	625	CTACGAAGCGGGGCGAGTGAACCTCCACAACGCCAGAA--CCAGSAGGCCATCTTTAACC	682				
QY	1850	ttaactgaccttcttctccgcgccaaggtctcctctgctcttgccttgcgcctgtgacacaaag	1909				
Db	683	TTTCACTGGCTCTTTCTTCCCGGCCAAGGTCTTCCCTGTCTCTTGGCGCGCTGGACCAAAAG	742				
QY	1910	gtctcgtgccttaagcttgacacacggtgggaataatgtggcgcgcggaaggtcgagcggaagc	1969				
Db	743	GTCTCTGGCTCTTAAGCTTGACACCGGGGGAATGGGGCGCGGAAAGGCCGCCACCGGGACA	802				
QY	1970	ggcggtcctcccgagatccag	1989				
Db	803	GGGCGGCCACCGGGCACACAG	822				

ACCESSION	mRNA sequence.
VERSION	BC386415
KEYWORDS	BC386415.1 GI:13279861
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

REFERENCE	1 (bases 1 to 952)
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: NIH Intramural Sequencing Center  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
row: b column: 12  
plate: LC3M308  
High quality sequence start: 7  
High quality sequence stop: 586.

FEATURES	Location/Qualifiers
source	1. .952

**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4583723"
/clone_lib="NH_MGC_15"
/tissue_type="adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pOTB7; Site_1: XhoI; Site_2
ECORI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the Laboratory of Gerald M. Rubin (University of
California, Berkeley) using 2AP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

```

Query Match	21.3%;	Score 555.4;	DB 11;	Length 952;
Best Local Similarity	86.8%;	Pred. No. 2.5e-105;		
Matches 723; Conservative	0;	Mismatches 81;	Indels 29;	Gaps 9

[illegible]



|||||  
Db 357 CATGCCCTGTGCCCCCTGGCCCGCTGGAGAG-AACCTACTGCGAGGACGCGAGAGA 415  
Qy 771 aatctactctctcttgagagaacagatgccgagttctcttggtcttgcagggagaggt 830  
Db 416 ACATCTACTTCTTC-TGCCAGAAAGCATGCCAGTTCCTGTGTCTTCAGGGAGGGGT 474  
Qy 831 ccacagccacagagcgcacacccgttgaggcttcctgagacgagagccatagccctaccggagt 890  
Db 475 CCCACGACACAGGCGACACCGTGGGGTTCCTGGACGAGG-CATTACGCCCTTACCGGGAT 533  
Qy 891 cgtctcaagagtcgacttggaagctctgagcagcggagagatgagatgtgtagatgttaa 950  
Db 534 CCTCTCAGAGTGCAGCTGAGAGCTGACGACGAGAGATGAGATGAGATGATGATGAT 593  
Qy 951 ggtctcaagagacccagaagcttcaagctgctgctgactgactgacgaacgaagcgaagccg 1010  
Db 594 GTCAAAAGA-----CCAACTTTCMACTGCTGTGACTCAGATCGAACAAGCAAGAGCGCG 648  
Qy 1011 tcaagggtgacacagctcctcttgagagctgacgaagcggagcggacgaacgaacgagatgtct 1070  
Db 649 TCAGGGGTGCACACAGCTCTTTCAGAGGCTGCAAGCGGGAGCTGCACGACAGCGATGTC- 707  
Qy 1071 cctgtcgtgcagagactgagtgtaacgtcgcgagatcacaagatttggaaggagagatga 1130  
Db 708 ----CTCCTGACGGCTAGCGCTGAGGAGCGCTGAGACAGAGATGAGGCGGAGAGGATCA 763  
Qy 1131 atatatcaaaaggtctctgaggaagatcaaccggcttgagaccccaagctcaag 1183  
Db 764 TTTT-----CAAAGGTTCTCGAAGACACCCCGGTTGAGCCCGCGTAAAG 811

RESULT 3  
AM969591 621 bp mRNA EST 01-JUN-2000  
LOCUS AM969591  
DEFINITION EST18168 MAGE resequences, MAGK Homo sapiens cDNA, mRNA sequence.  
ACCESSION AM969591  
VERSION AM969591.1 GI:8159435  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 621)  
Hedge, P., Qi, R., Abenathy, K., Dharap, S., Gaspard, R., Gay, C., Holt  
I., E., Saeed, A. I., Sharov, V., Lee, N. H., Yeaman, T. J. and  
Quackenbush, J.  
TITLE Assessment of gene expression patterns in a model of colon tumor  
metastasis using a 19,200 element cDNA microarray  
JOURNAL Unpublished (2000)  
COMMENT Contact: John Quackenbush  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 3528  
Fax: 301 838 0208  
Email: johnq@ligr.org  
Plate: 267  
Seq primer: Forward.  
Location/Qualifiers  
1. 621

FEATURES  
Source  
1. 621  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="MAGE resequences, MAGK"  
/note="Vector: pBluescriptSKm"  
BASE COUNT 165 a 155 c 144 g 157 t  
ORIGIN

Query Match 16.7%; Score 437; DB 10; Length 621;  
Best Local Similarity 97.5%; Pred. No. 9.8e-81;  
Matches 465; Conservative 0; Mismatches 10; Indels 2; Gaps 2;  
Qy 2134 tgacgtctgagctctctgacactcagagtgcagaccagacggtcgtgctgctgc 2193

|||||  
Db 619 TGACGTTTGGCATTATATATGCCCTCAGAGTGCAGAACTCAGACGGCTTCGGCTGAGCC 560  
Qy 2194 tagggcaaacgacacactagagaccgcgcgcctcttgagggaataacataaagagagac 2253  
Db 559 TAGGGCAACAGCCCAACCTTAGAGCCAG-CGGGCTTTCGGGGAATACTAAAGAGAGAC 501  
Qy 2254 atctaaatgtaatatgttaactgtttcaagataattactcttgagggaataaaggtttt 2313  
Db 500 ATCTAAATGTATGTATTAACTGTTTCAAGATATATTATCTTGGGAAATATAGGCTTTT 441  
Qy 2314 gcttgacttgacactaatttgtaacgttaacttcgtaacttgacacacacactgaagatgac 2373  
Db 440 GCTGACCTTGCACTAATTTGTACAGTTAAGTTCGTAATTTGACACACACCTGAAGATGCC 381  
Qy 2374 tccaccttgtagggtctagggcctttatataagcccttggtgagacccagggcccttc 2433  
Db 380 TCCACCTTTGTAGGGGCTTGTAGGGCTTTTATCAGCCCTGGGTGAGACCCAGGGCCCTTTC 321  
Qy 2434 ctctccctccctctctggtcattctctgacttgtagagaaatgcttaagaagtgtga 2493  
Db 320 CTTCCTTCCCTTCTGTCATTTCTCTGACTTGTAGAAATGTCCTAAGAAAGTGTGA 261  
Qy 2494 ctacagaccttgatgattccatggtccaaatlagcgtgtagtgagactgagagaagactta 2553  
Db 260 CTCACAGACCTCTGATGATTCATGTGTCATTAAGCGCTGATGGAGCTGAGAAAGCTTA 201  
Qy 2554 aatcaatggatctgctctggttggtggaatattagggccgagatgctgagagagt 2610  
Db 200 AATCCAAATGGGATC-TGCCTGTGTTGGCAATTTAGGGCCGAGATGCTCGAGGAGT 145

RESULT 4  
A1732648 666 bp mRNA EST 13-DEC-1999  
LOCUS A1732648  
DEFINITION nm73f01.x5 NCI\_CGAP Co9 Homo sapiens cDNA clone IMAGE:1073881 3'  
similar to TR:Q12899 Q12899 ACID FINGER PROTEIN.;, mRNA sequence.  
ACCESSION A1732648  
VERSION A1732648.1 GI:5053761  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 666)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP).  
JOURNAL Unpublished (1997)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgaphs-remail.nih.gov  
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,  
Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/ILNL at:  
[www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

FEATURES  
Source  
1. 666  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1073881"  
This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: National Cancer Institute, Cancer Genome  
Anatomy Project (CGAP), Tumor Gene Index  
This read has been verified (found to hit its original self in the  
correct orientation)  
Insert Length: 1309 Std Error: 0.00  
Seq primer: -40up from Gibco  
High quality sequence stop: 462.  
Location/Qualifiers  
1. 666  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1073881"



```

187 ATTATCTAGAGCCCTCAGCTCCCTGAGCTCCCTGAGCCCTCCCTGAGCCCTGAGCCCTC 128
QY 2150 tctgcaactcagagtgacgaacacagacggttcgtgctgctcctgaggaacagccaac 2209
Db 127 TCTGTACTCAGACAGACGAGAACACAGAGGCTTGGCTGTCTCAGGCGCAACACCCAAAC 68
QY 2210 ctgaagaacccgagccttcgagggaacaaactaagaagagagactcctaaatgtaagt 2269
Db 67 CTAGAGACGACGAGGCTTTCCGGGGAANA--AAAGAAAAGACATCTAAATTAATGT 11
QY 2270 ttaactgt 2279
Db 10 TTAACGTGT 1

RESULT 6
AI222799 525 bp mRNA EST 02-FEB-1999
LOCUS gp39b10.x1 NCI_CGAP_C08 Homo sapiens cDNA clone IMAGE:1925371 3',
DEFINITION mRNA sequence.
ACCESSION AI222799.1 GI:3805002
VERSION AI222799.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 525)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert length: 1145 Std Error: 0.00
Seq primer: -40bp from Gibco
High quality sequence stop: 451.
FEATURES
Source
Location/Qualifiers
1..525
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1925371"
/clone_lib="NCI_CGAP_C08"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dt) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 106 a 151 c 157 g 111 t
ORIGIN
Query Match 13.7%; Score 358.6; DB 10; Length 525;
Best Local Similarity 89.1%; Pred. No. 1.9e-64;
Matches 480; Conservative 0; Mismatches 44; Indels 15; Gaps 8;
QY 1748 ccgctgagcagatcccgacgagcggtgagatcgccctgacatcagagcgggcagct 1807
Db 524 CCGGTGAGCGAGATCCCGCGC-GGCGTGAGAGTCCGCTGAGACTAGAGCGGCGAGGT 466

```

```

QY 1808 gaccctccacaagcccaagagcccaagggccacatcctcaactcactgctcttc 1867
Db 465 GA-CCTCCACAGCGCCAG--CCCAAGAGCCCATCTTCACTTACGCTCTT--CT 411
QY 1868 cggccaaggtctcctcctgctccttgccgccttgacacaagaaggtcctgactga 1927
Db 410 CCGGCAAGGTCTTCCCTTCTTCTTGGCTGTG--AAAAAGTTCTCTTACGCTGA 354
QY 1928 caaggaggaaatgagcgcgaggaagcgagcgagcgagcgagcgagcgagctc 1987
Db 353 -AAGCTGAACTGAGGCGCGCGCAAGGCGCGCAAGCGGAGCGGCGCTCCGGATCC 295
QY 1988 agctcgcctcggcgaagtgtcgccgcggggggtcccgctgcggcgaggagaga 2047
Db 294 AGCTCGCCCTGCGCAGTGTGCGGCGCGGCGGCTCCCTGCTCCGCGTGAAGCAGAGA 235
QY 2048 aacacggagactgtgctcgcagacagcggttcttactatattatctagaccctca 2107
Db 234 A--CAGGGGACTTGAAGTCTCGAAGCGGCTTTTACTTATTATTACTTAGGCCCTCA 177
QY 2108 gctcctcagctcgtcgtgagcctcctcgtgagcgtctgacctgacctgaagtga 2167
Db 176 GCTCCCTGACGTCTGAGGCTCCTGTGAGCTGTGCGCTTCTGCACTCAGAGTGCA 117
QY 2168 gaaccacagagcgttcgtgctgctagggacaagccaactaggaaacccgagcct 2227
Db 116 GAACCCACAGCGGCTTGGCTGTGCTAGGCAACGCCAACCTAGAGCCAGCGGCTT 57
QY 2228 ttcggggaaaaactaagaagagacatctaaatgtatgttaactgttcaagt 2286
Db 56 TCGGGGAAAA--AAAGAAAAGACATCTAAATGTTTAACTTTCAAAAT 1

RESULT 7
AM820993 450 bp mRNA EST 17-MAY-2000
LOCUS RC2-ST0301-240300-016-d01 ST0301 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM820993
ACCESSION AM820993.1 GI:7913987
VERSION AM820993.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 450)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.G.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?lpl=st2-RC2-ST0301-240
300-016-d01&t3=2000-03-24&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 36
High quality sequence stop: 450.
FEATURES
Source
Location/Qualifiers
1..450

```

BASE COUNT	87 a	148 c	109 g	106 t
ORIGIN	low stringency conditions."			
Query Match	13.4%; Score 351; DB 10; Length 450;			
Best Local Similarity	91.8%; Pred. No. 7,1e-63;			
Matches	416;	Conservative	0;	Mismatches 30; Indels 7; Gaps 4;
OY	663	gcccaaggagtagcagagcagaagactccatgtgcccctgtgccccttgagcccgctggag	742	
DB	450	GACACAGAGAGAGACAGTCAGACACTCCCATGGCCCTGCTGCTGGGCCCGCTGGAG	391	
OY	743	ataacttactgtggaagagacagcgcggaagaatctacttcttcttcgcgaaacgatgcga	802	
DB	350	ATAACTTACTGTCGAGAGACGCGGAGAAAGATCTACTCTTTC-TTCGAGAAAGATGCCGA	332	
OY	803	gtctctctgtgtgtctctgcagaggaggtgccacgacacagcgacacacgctgggttctc	862	
DB	331	GTCTCTCTGTGTGTCTCTGAGGAGGAGGTCCACGACACGACGCGCACACCGTGGGTTCT	272	
OY	863	ggagcagagccatcagccctaccgagatcgtctcagaagtcgactgtgaagctctgaagac	922	
DB	271	GGACAGAGGCCAATTCAGACCCCTTACCGGAGTCTCTCAGAGTGCACATCGAAGCTCTAGACAC	212	
OY	923	ggaagagatgtgagatctgttagagatgtcaaaagtgtcagaagaagaccagaagctcgaagtgcgc	982	
DB	211	GGAGAGAGATGATGATTTGATGATTTGAAAGTGTCAAGAAAGACCAAGAAAGCTTCAAGTGGCGC	152	
OY	983	tgaactcagatcgaaacaagaagacgcgtcgaaggtgtcacacacagctccttgaagagcttga	1042	
DB	151	TGACTTCAGATCG--CAACGACAAAGACATCAGCTGGTAACA--GCTTTGAGAGGCTGC-	96	
OY	1043	agcggagactgtgagcagcagcagatgtctcgtctgctggcagagactgagtggagactcgg	1102	
DB	97	AGCAGAGACTGTGTGACGACGCGCATCTCTCTGTCGCGAGGACTGATGTGATCGCTGG	38	
OY	1103	agtcacagatttgaagagagagatgaatga	1135	
DB	37	AGTCACAGATTTGGCAGAGTGAAGGACGTGTANA	5	
RESULT	8			
AK007445				
LOCUS	AK007445 1779 bp mRNA HTC 05-JUN-2001			
DEFINITION	Mus musculus 10 day old male pancreas cDNA, RIKEN full-length			
ACCESSION	enriched library, clone:1810012B10, full insert sequence.			
VERSION	AK007445.1 GI:12841001			
KEYWORDS	CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) 10 day old male pancreas cDNA to			
	mRNA, clone:11b-RIKEN full-length enriched mouse cDNA library			
	clone:1810012B10.			
ORGANISM	Mus musculus			
	Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 1779)			
AUTHORS	Carninci, P. and Hayashizaki, Y.			
TITLE	High-efficiency full-length cDNA cloning			
JOURNAL	Methods in enzymology. 303, 19-44 (1999)			
MEDLINE	99279253			
PUBMED	10349636			
REFERENCE	2 (bases 1 to 1779)			

AUTHORS	Carinci, P., Shibata, K., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	genome research. 10 (10), 1617-1630 (2000)
PUBLISHED	20493974
REFERENCE	11042159
AUTHORS	3 (bases 1 to 1779) Shibata, K., Itoh, M., Aizawa, K., Nagoka, S., Sasaki, N., Carinci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, K., Iwawa, M., Ohara, E., Matsumi, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome research. 10 (11), 1757-1771 (2000)
PUBLISHED	20530913
REFERENCE	11076861
AUTHORS	4 (bases 1 to 1779) The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
PUBLISHED	5 (bases 1 to 1779)
REFERENCE	Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arakawa, T., Carinci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirokawa, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kato, H., Kikawa, J., Kojima, Y., Kono, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT	Please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and genome science laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGAGAGAGAGATCCAGACGCTCTTTTATTTTATTTTNN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Rot - 20.0. Second strand cDNA was prepared with the primer adapter of sequencel5' GAGAGAGAGAGCGCCGCAATTATTTCTCGAGTTATTAATTATATCCCCCC 3'. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.
FEATURES	Location/Qualifiers
SOURCE	1. 1779 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /db_xref="MGI:1901147" /db_xref="MGI:1916347" /clone="1810012B10" /sex="male" /tissue_type="pancreas" /clone_id="RIKEN full-length enriched mouse cDNA library" /dev_stage="10 day old"
BASE COUNT	481 a 447 c 492 g 359 t

## ORIGIN

```

Query Match      13.2%: Score 343.4; DB 12; Length 1779;
Best Local Similarity 69.2%: Pred. No. 3.5e-61;
Matches 683; Conservative 0; Mismatches 261; Indels 43; Gaps 14;

Qy 419 tgaagtacatcgaattcaaggaaaggaacacgcgtgcgtgag-ggaaccggagtga 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 31 TAAAGTACTACTGATTGAGGAAAGAAATTGGAGTGGACGAGTCAAGAAACTGTGGA 90

Qy 478 cggagctgggaaggaacacgtgacccgcgaaccccgctccctggaaaggggtgcacatg 537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 CAGGACGGGGAAGGACGGTATGCCCTCAACCCCGTCCCTCG-----GGACATTC 142

Qy 538 agctgcctgcgtacccctcgtgcggggcgctggaagatgaggtgcacatccctgtg 597
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 ACCAAGTAGCCTGTCTCTGACACCCCACTTCGAGATGCGGTGACACAGCCTCG 202

Qy 598 gacacacctctgcgggtctgcgcctcccgcgctcccgatggggcccaatccctgt 657
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 GACACACGTGTGCTTACTCTGCTCC---CGCGACCCCAATGGAGCCCACTGCTGT 259

Qy 658 ggcgaagatcctgctcgcgcctcgcgaagagagtagcagcgcaagactccatgccc 717
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 260 G-----CCCTCTCTGACAGGGGCGAGAGAGAAATCCAGGCCCGCACTGGCC 309

Qy 718 cctgtgccccggggcgccgctgggagataactactgtggaagacagcgcgagaagatcta 777
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 310 CCGGTGTCCCTAGGTCCCTTAGCGA-GACCTACTGTAGAGCACGAGCAATMAATCTA 368

Qy 778 cctcctctcgagagaacatgccgagctcctcgtgtgtctgcaggaaggggtccacgc 837
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 TTTCTCTC--GTAGACACGACGGGACCTGCTGCTTCTGCGAGGAGGCCCCGCC 426

Qy 838 accagcgacacacccgtgggttctcctgcgaagagccatcaccctaccggagatcctca 897
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 ACCAGGCGACACACTGTGGGTTCTCGATGAAGCCATCCAACTCAGAGGATGCTCA 486

Qy 898 ggaatcgactggaagctctgcgaagagagatgagatgtgaagtgtcaaa 957
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 GGAGTCGGTTGAAGCTTAAAGATGAGAGGAGGACAAATAG-AAAGATGAGAAATGTCAA 545

Qy 958 gaagacaggaagcttcaagtcgagctgagctcagatcgacaagaagccgctcaggt 1017
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 GAACACCGAAGCTTCCAGAGAGTTTGATGCAAGTTGA-----AAGCAAAAACATCAG 599

Qy 1018 gcacacagctccttgaaagctgcgaagcgagctgcagcagcagcagatgtctcgtg 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 GTAGAACCTGCTTTTGAAGGCTGAAGGGGGAGTTGTGACACAGAGATGCTCTGCTG 659

Qy 1078 ggcgagagctgagtggtcgtcgtcgagatcacaagatttgaagagagagatgataatc 1137
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 660 A-----ACAGGCTGGAAGAACTAAGCAAGATCTCAAGAGAGAGAGATACATC 713

Qy 1138 acaaaagctcctggaagagtaaccgggttgagcccgacgtcgaagagctcgagagga 1197
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 AGCAAAAGTCTCCAGAGAGTGCAGCCGC--TGAGTACCCAGGTGGAGAGACT--AGAGGAGA 771

Qy 1198 agtgtcaagcagcagcaagtgaagcttcaagaatgclcagaagtcacagcagcagtgct 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 AGTGTACAGCACCGACAGTAGTGTGCAAGATGACCACTCAA--TCAGAGCAGGTAC 830

Qy 1258 gaagtgaagacttttgaagctcgaagcatttctccctgacccgttccaagaagatccg 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 GAGCTGAAGACTTTTGTAGGCCAGAGGCCATTCCT--CTGATCTTTGTCAAGAAATCCG 889

Qy 1318 tgattccacagaaataactcaacctccagagatgaatgaatgttctcaagaact 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 890 AACACTCCACAGGAAATACATTAGCTCCCAAGATGATGAGACATTTTC--AGAAACT 948

Qy 1378 tggcgatcatctggaataagattcag 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 949 TGATGACACACTCGAGACAGATTCAG 975

```

RESULT 9
A1821412 429 bp mRNA EST 09-JUN-1999
LOCUS ne58d09.x5 NCI_CGAP_C03 Homo sapiens cdna clone IMAGE:901553 3',
DEFINITION mRNA sequence.
ACCESSION A1821412
VERSION A1821412.1 GI:5440491
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 429)
AUTHORS NCI/NIDR-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bhrp/image/image.html

```

This read is a RESEQUENCE of a previously sequenced human clone  
Original clone citation: National Cancer Institute, Cancer Genome  
Anatomy Project (CGAP), Tumor Gene Index  
This read has been verified (found to hit its original self in the  
correct orientation)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 424.

## FEATURES

## Source

Location/Qualifiers  
1..429

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:901553"  
/clone\_lib="NCI\_CGAP\_C03"  
/sex="pooled"  
/tissue\_type="colon"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
was prepared from 12 pooled bulk tumor samples and primed  
with a Not I - oligo(dT) primer. Double-stranded cDNA was  
ligated to Eco RI adaptors (Pharmacia), digested with Not  
I and cloned into the Not I and Eco RI sites of the  
modified pT73 vector. Library went through one round of  
normalization."

BASE COUNT 95 a 121 c 125 g 88 t

## ORIGIN

```

Query Match      12.5%: Score 325; DB 10; Length 429;
Best Local Similarity 91.3%: Pred. No. 1.8e-57;
Matches 390; Conservative 0; Mismatches 30; Indels 7; Gaps 4;

Qy 1903 acaaaagctcctgacatgaagctgacacggggaatgaggcgcgagagggcgagaag 1962
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10 AAAAAGGTTCTCTGCTTACAGCTGA-AAAGCTGAAGTGGGGCGCGAAGGCGCGAAG 68

Qy 1963 cggaaagcgagctcctcggaatcaagctcggcccttgcaagtgtagccgggggact 2022
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 CGGAGACGGCGGCTCTCGGGAGTCCAGCTCCGCCCTGACAGTGTGCGCCGGGGCT 128

Qy 2023 cccctgcccgcgtgaggcgagaaacacgagagactgagctcgaacacgcgtgttt 2082
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      129  CCCTGTGCCCGCTGAGCGCAGAGAA--CA6GGGACTTGAATCTCGAACA6CGGCTGTWT 186
QY      2083  ttacttattatcttagagccctcagctcctcagctcctcagagctcctcgttagagctct 2142
Db      187  TTACTTATTATTATCTTAGGCGCCTCAGCTCCCTGACGTCCTGAGCCTCCCTGTCAGCCTCT 246
QY      2143  ggcctctctgcagctcagagtgacagacacacagacagctcgtctgcttagggcaac 2202
Db      247  GGCTTCTGACACCTCAGAGTGACAGAACACACACGCGCTTCGCGTGTGCTAGGCGAAC 306
QY      2203  agccaactaggaaccccgccgctctcgggggaaaaactaagaagagacatctaaat 2262
Db      307  AGCAACCTTAGGAACCCGCGCCTTCGGGGAACACAAAGAAAGAGAGCTCTAATC 366
QY      2263  gtaatgttaaacgtctcaagataatcctctgggaaaaatcagggcttctgctgagct 2322
Db      367  AAACGTG---CAACTGTTCAGATATATCTTGGGAAACAT-AGGGTTTGTGCTGACTT 422
QY      2323  gcaactaa 2329
Db      423  GCACATA 429

RESULT  10
LOCUS   AM610230 517 bp mRNA EST 23-MAR-2000
DEFINITION RC2-ST0301-300100-013-h02 ST0301 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM610230
VERSION   AM610230.1 GI:7314971
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 517)
AUTHORS   HCGP http://www.ludwig.org.br/ORESTES.
TITLE     The FAPESP/LICR Human Cancer Genome Project
JOURNAL   Unpublished (1999)
COMMENT   Contact: Simpson A.J.G.
           Laboratory of Cancer Genetics
           Ludwig Institute for Cancer Research
           Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
           Brazil
           Tel: +55-11-2704922
           Fax: +55-11-2707001
           Email: asimpson@ludwig.org.br
           This sequence was derived from the FAPESP/LICR Human Cancer Genome
           Project. This entry can be seen in the following URL
           (http://www.ludwig.org.br/scripts/gethtml2.pl?L-RC2&t2-RC2-ST0301-
           300100-013-h02&t3-2000-01-30&t4-1)
           Seq primer: puc 18 forward
           High quality sequence start: 45
           High quality sequence stop: 517.
FEATURES
         source
           1..517
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone_lib="ST0301"
             /dev_stage="Adult"
             /note="Organ: stomach; Vector: puc18; Site:1: SmaI;
             Site:2: SmaI; A mini-library was made by cloning products
             derived from ORESTES PCR (U.S. Letters Patent Application
             No. 196,716 - Ludwig Institute for Cancer Research)
             profiles into the puc 18 vector. Reverse transcription of
             tissue mRNA and cDNA amplification were performed under
             low stringency conditions."
BASE COUNT      92 a 172 c 128 g 125 t
ORIGIN

```

```

Query Match      12.38; Score 320.6; DB 10; Length 517;
Best Local Similarity 88.28; Pred. No. 1.5e-56;
Matches 442; Conservative 0; Mismatches 44; Indels 15; Gaps 8;

```

```

QY      683  gccacagagatagcagcagagagaccatctgagccctgtgccccctggccgcgtggag 742
Db      517  GCCAAGAGAGAGACAGACAGACAGACTCCATGGCCCTGTGCCCCCTGGCCGTGGAG 458
QY      743  ataacttactgcagagagcagcgagaaagatcactctctctgcagaaagatgcga 802
Db      457  A-AACTTACTGCGAGGACAGCGCGAGAMAGATCTACTTCTTC-TGCCAGAAAGATGCCGA 400
QY      803  gtctctctgtgtctctgcagggaggtcccaacgacccaagcgagacacgttggttct 862
Db      399  GTTCTCTGTGTCTTGTGCAAGGAGGGTCCACTGACACAGCGACACCGGGGTTCT 340
QY      863  ggaagagacatcaagccctcaccggatctctcaggaagtcagatggaactcgaac 922
Db      339  GGAGAGGCCCATTCAGCCCTACCGGATCTGCTCAGAGATCGATGAGACTCGACAC 280
QY      923  ggaagagatgagatgtgtaagatgaagtgtcaagaagacccaagctcaagtgcgc 982
Db      279  GGAGAGAGATGAGATGTG-AGGATGTAAAGTGTCAAGAAAGACCAAGAACTTCAAGTCTGC 221
QY      983  tgactcagatcgaaacaagcagaagccgtcaagggtgcacacagctcttaagagctgca 1042
Db      220  TGACTCAGATCGTATCCTTAGAAGCATC---AGGTGAGACAGCT-TTTGAGAGGCTGC- 165
QY      1043  agcggagagctgcagcagcagatgtctcctctgctgcgcaggaagtgtgacgtcgg 1102
Db      166  AGCAGAGAGCTGAGACACACGATGTCTCTGCTGGCCAGGCTGAGGAGACTGGAGACG 107
QY      1103  agtcacagattggaagagagagatgaatataatcacaaagtctctgaaagtacccc 1162
Db      106  AGAC-----TTTGGAAAGAGAGGATGAATATATCACAAAGCTCTGAGGAAGTCAACC 52
QY      1163  ggccttgagccccaagcgcgaag 1183
Db      51  GGCTT-GAGCCCGACGTCAGG 32

RESULT  11
LOCUS   BF431333 449 bp mRNA EST 29-NOV-2000
DEFINITION naa43f01.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3259152 3'
ACCESSION BF431333
VERSION   BF431333.1 GI:11443447
KEYWORDS EST.
SOURCE   human.
ORGANISM Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 449)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer genome Anatomy Project (CGAP),
           Tumor Gene Index
JOURNAL   Unpublished (1997)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgaops-remail.nih.gov
           Tissue procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
           Emmert-Buck, M.D., Ph.D.
           cDNA Library Preparation: M. Bento Soares, Ph.D.
           cDNA Library Arrayed by: Greg Lennon, Ph.D.
           DNA Sequencing by: Washington University Genome Sequencing Center
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL, send email to:
           info@image.llnl.gov
           Seq primer: -40UP from Gibco
           High quality sequence stop: 288.
FEATURES
         source
           1..449
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:3259152"
             /clone_lib="NCI-CGAP_Kid11"

```



Source	Organism	Human
REFERENCE 1 (bases 1 to 418)	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	
JOURNAL	Unpublished (1997)	
COMMENT	Contact: Robert Strausberg, Ph.D. Email: gcgaps@email.nih.gov Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ULML at: <a href="http://www-bio.lnli.gov/bbrp/image/image.html">www-bio.lnli.gov/bbrp/image/image.html</a> Insert Length: 1123 Std Error: 0.00 Seq primer: -400P from Glbco High quality sequence stop: 404. Location/Qualifiers 1. 418 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1873192" /clone_1ib="NCI_CGAP_Co8" /tissue_type="adenocarcinoma" /lab_host="DH10B" /note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from colon adenocarcinoma, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bernaldo."	
BASE COUNT	92 a 118 c 117 g 91 t	
ORIGIN		
Query Match	11.8%; Score 308.2; DB 10; Length 418;	
Best Local Similarity	89.0%; Pred. No. 5.6e-54;	
Matches 380; Conservative	0; Mismatches 38; Indels 9; Gaps 4;	
QY 1860	ctttctccggcccaagcttcctccctgctcttcggccgctgacacaaggtgctgacct 1919	
DB 418	ctcttcttccggcccaagcttcctccctgctcttcggccgctgacacaaggtgctgacct 362	
QY 1920	tagagctgacacggggaaatggtggcgcgcgagagggcgagcgagagacggcgctctc 1979	
DB 361	TACGCTGA -AAGGCTGAAGTGGGGCGCGCGAAGGCGGCGAAGGCGGAGACGGGCTCTC 303	
QY 1980	cggagctcagctcgcccttgccagtggtgcggccggggggtccctgtgcccgctgag 2039	
DB 302	CGGGATCCAGCTCCGCCCCCTCGGCGAGTGTGGGGCGCGCGGCGCTCCCTGCGCGCGTAG 243	
QY 2040	gcgagagaagaacaggggagcttgagcttgacaacagaggtgttttaactattatctta 2099	
DB 242	GCGAGAGAA -CAGGGGACTTGAGCTCTCAACACAGGGGTTGTTTACTTTATTTATCTTA 185	
QY 2100	ggacctcagctccctcgaagctctgagctccctcctgagacgctcgtgctctcctgacct 2159	
DB 184	GGCCCTACGCTCCCTGACGCTCTGAGCTCCCTCTGTGACGCTCTGGCCCTTCTGTGACCTC 125	
QY 2160	agagtgacagaacacagacggcttcgagctgtgcttaaggcaacgacacactlaggaacc 2219	
DB 124	AGAGTGACAGAACACAGACGGCTTGCGCTGTGCTTAGGGCAACAGCCAAACCTAGAGAGCA 65	
QY 2220	gcgggctcttcgggggaaactaaagaagagacactaaatgtaactgtaactgtt 2279	
DB 64	GCGGGCTTTTCGGGGAAAAA - - - - - AAAAGAAAAAGACACTCTTAATAATGTTAAACTGT 8	

Oy	2280	tcaagat	2286
Db	7	TCGAAT	1
RESULT	14		
LOCUS	A1791944		
DEFINITION	nm73f01.y5 NCI-CGAP Co9 Homo sapiens cDNA clone IMAGE:1073881 5' similar to TR:Q62157 Q62157 ZINC FINGER PROTEIN ;, mRNA sequence.		
ACCESSION	A1791944		
VERSION	A1791944.1	GI:5339660	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 491)		
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index		
COMMENT	Unpublished (1997) Other_ESTS: nm73f01.x5 Contact: Robert Strausberg, Ph.D. Email: cgapbs@mail.nih.gov Tissue Procurement: Ian Kirsch, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by: Greg Lennon, Ph.D. DNA sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LTML at: <a href="http://www-bio.jnl.gov/bdrp/image/image.html">www-bio.jnl.gov/bdrp/image/image.html</a>		
FEATURES	This read is a RESEQUENCE of a previously sequenced human clone Original clone citation: see original entry for original citation Information This 5' resequenced clone has no previous 5' data to verify this new read against Insert length: 1309 Std Error: 0.00 Seq primer: -40RP from Gibco High quality sequence stop: 462. Location/Qualifiers 1..491 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1073881" /clone_1bp="NCI-CGAP Co9" /tissue_type="colon tumor RER+" /lab_host="DH10B" /note="Organ: colon; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from RER-colon tumor, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector. Library is not normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo (Soares4 )"		
BASE COUNT	135 a	115 c	149 g
ORIGIN	92 t		
Query Match	11.7%	Score 306;	DB 10:
Best Local Similarity	88.8%;	Pred. NO. 1.7e-53;	Length 491;
Matches 444;	Conservative 0;	Mismatches 20;	Indels 36;
Gaps 9;			
Oy	1030	ttgagaagctccaagcggaagctgcgaagcagcatctccctgtcgagcaggactga	1089
Db	28	TTGAGAGGCTGC AGCAGACTGTGAGCAGCAGCATGTCTCTCCGTGGCCAGGCTGAGG	86
Oy	1090	gttgtaaccttcggagtcacagatttgaaagagaggaaltgaatatatcacaaagtctct	1149





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 13:07:47 ; Search time 519.52 Seconds

(without alignments)  
4307.089 Million cell updates/sec

Title: US-09-867-034-5

Perfect score: 2610  
Sequence: 1 gatgtgcgcagcctccagag.....ccgagatggtcgcagagagt 2610

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_1101:\*

1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	208	8.0	3227	22	AA158226 Human polynucleoti
2	192.2	7.4	2351	22	AA160012 Human polynucleoti
3	76.4	2.9	520	22	AA123948 Probe #13881 for g
4	76.4	2.9	520	22	AA149253 Probe #17939 used t
5	76.4	2.9	520	22	AA109542 Probe #9533 used t
6	76.4	2.9	3518	19	AAV57905 Bovine butyrophilin
7	69.2	2.7	2981	21	AACT7707 Human cancer assoc
8	67.6	2.6	43226	20	AAK60263 Nucleic acid seque
9	63	2.4	712	21	AAAT8067 CDNA encoding huma
10	63	2.4	712	22	AAAT8805 Colon tumour relat
11	63	2.4	43226	20	AAK60263 Nucleic acid seque

12	60.6	2.3	376	22	AAH72187
13	60.6	2.3	509	22	AAH71127
14	60.6	2.3	750	22	AAH72832
15	60.6	2.3	14180	19	AAV05159
16	59.6	2.3	571	20	AAV87250
17	59	2.3	3479	22	AAAS21305
18	58.8	2.3	3161	16	AAO80233
19	58.8	2.3	3344	16	AAO80228
20	58.8	2.3	5191	21	AAZ8617
21	58.2	2.2	2280	17	AAI36127
22	57.6	2.2	400	22	AAE64260
23	57.6	2.2	411	22	AAE65013
24	57.4	2.2	243	21	AAH30668
25	57.2	2.2	395	22	AAE64375
26	57	2.2	428	22	AAE64255
27	57	2.2	3515	22	AAE75343
28	57	2.2	5403	20	AAV84140
29	56.8	2.2	387	22	AAE67194
30	56.8	2.2	2904	19	AAV57904
31	56.4	2.2	412	22	AAE67777
32	56.2	2.2	471	22	AAE69324
33	56.2	2.2	49999	20	AAZ23899
34	56	2.1	558	22	AAI16803
35	56	2.1	558	22	AAI39486
36	56	2.1	558	22	AAI40381
37	56	2.1	4122	22	AAAD10239
38	56	2.1	7715	19	AAVA3042
39	55.8	2.1	383	22	AAE64661
40	55.8	2.1	1462	20	AAZ10861
41	55.8	2.1	2095	22	AAAD07708
42	55.6	2.1	386	22	AAE64628
43	55.6	2.1	402	22	AAE64814
44	55.6	2.1	700	11	AAO05263
45	55.6	2.1	700	15	AAQ57505

#### ALIGNMENTS

RESULT 1	
AA158226	AA158226 standard; cDNA: 3227 BP.
XX	XX
AC	AA158226:
XX	XX
DT	22-OCT-2001 (first entry)
XX	XX
DE	Human polynucleotide SEQ ID NO 429.
XX	XX
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW	leukemia; ss.
XX	XX
OS	Homo sapiens.
XX	XX
PN	WO20015312-A1.
XX	XX
PD	26-JUL-2001.
XX	XX
PF	26-DEC-2000; 2000WO-US34263.
XX	XX
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662192.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	XX

Human cervical can  
Human cervical can  
Human cervical can  
Mouse butyrophilin  
EST clone B0260.  
Human cDNA sequenc  
Rat NDF clone 42B  
Rat NDF clone 22 D  
RPP1-Wsa genomic n  
Mouse neuropeptide  
Novel human polynu  
Novel human polynu  
Human colon cancer  
Novel human polynu  
Novel human polynu  
Human rGF-beta rec  
Mouse prothrombina  
Novel human polynu  
Bovine butyrophilin  
Novel human polynu  
Human cervical can  
Human LOBO homology  
Probe #6736 for ge  
Probe #8172 used t  
Probe #9067 used t  
Mouse long whey ac  
Mus musculus Cgamm  
Novel human polynu  
Hyaluronate synth  
Human secreted pro  
Novel human polynu  
Novel human polynu  
Sequence of neuron  
Rat GAP-43 promote



PR	09-JUL-2000;	2000US-0598042.
PR	19-JUL-2000;	2000US-0628312.
PR	03-AUG-2000;	2000US-0653450.
PR	14-SEP-2000;	2000US-0662191.
PR	19-OCT-2000;	2000US-0693036.
PR	29-NOV-2000;	2000US-0727344.
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	Zhao QA, Zhou P, Goodrich R, Demanac RT;	
XX		
DR	WPI: 2001-442253/47.	
DR	P-PSDB; AAM40856.	
XX		
PT	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	such as central nervous system injuries -	
XX		
XS	Claim 1; SEQ ID NO 4001; 10078pp; English.	

CC The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAM442213) with neurotrophic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX  
S0 Sequence 2351 BP; 577 A; 575 C; 680 G; 518 T; 1 other;

Query Match 7 4% Score 192.2; DB 22; Length 2351;  
Best Local Similarity 56.1%; Pred. No. 3.8e-39;  
Matches 667; Conservative 0; Mismatches 493; Indels 29; Gaps 15

OY	715	gcccctgtgcccctggtgcccgcgtcgttggaagaaactaactactctgagagagacacggtcgagaagaat	774
Db	271	gctggtgtccacaacactggtgttttgagagagagatgtctgtccaaagacagagagaagaat	330
OY	775	ctactctctcttgagagaacgattgcgcgaattcctctctgtgtctgtcttcgaagagaggtgccca	834
Db	331	ctactctctc-tgtagagatgataagatgcagttgtgcgtgtgtgtgttcctcgagagagctgtggg	389
OY	835	cgcacagcgagacacgcgcgtgggtcttcctctgaaagaaagc-cattcagccctaccgagatcgt	893
Db	390	agcacgtaccacccacacccatgcgtctccctcgtgagagatgcagcgggtctccctccataaggaaca	449
OY	894	ctcagagatgcgaactgaaagctctctgacacagagagagagatgaaatgttaagatgttaaaagt	953
Db	450	atccataagtgctcttaaatgtcttaataaagagagagagagatc-caagaatccaaagtc	508
OY	954	tcaagaagaaccagaagcttcaagtgcggtcgtacatcgatctgaacaaagaagaagccgtcaa	1013
Db	509	aagagaataataaagaatgcagctctctcgtactcagctcagtgctccaccaagaagacaacagtt	568
OY	1014	gggtgcacaacagctcctcttgagagagctctgcacaagcgggagatgcgcagcgcagcagatgttctct	1073
Db	569	gattctctga-----gttcgcacacctgtgggaagtcttctagaaggaacacagagacatcct	622
OY	1074	gctggcgcagagactgagtgtgtagcgtctcggagctcaacagatttgtaagagagagagataata	1133
Db	623	cttaagcaca-----atgtgagagccacaggaatgtgggacaactctgagagcaacgcggatgaatt	676

QY	1134	tatccacaaggtctctctgaggaagtcacccggtcttgagaccacgtacaagaagctcgag	1193
Db	677	tgaattgtctgtctgtctgagagatctgcggttttagtctctttaattgaagaact--ggag	734
QY	1194	gaggaagctgcaacgaacccgaagtgagctcttacaagaatgcacagaatcaagccgagcag	1253
Db	735	gggaagaatgaaagagccacgaagagggatgcttcgacgagacacagaagaactcta-ataag	793
QY	1254	gtctgagatgaaagacttttgtgagctcgtgaagagcaattctctccgtgacctgttcaagaaga	1313
Db	794	atgtgaacacagaagaagtgcggaaaccggtgtgctgtgtc--gccagagcttggccagagga	852
QY	1314	tccgtgatttccacaaggaataatactcacctccaccagagatgatatgaagtgtctcaagaa	1373
Db	853	tctcggaactttccccaagcagagccctcccgctgcagagagagatgaagaagttt--cttga	911
QY	1374	aacttgagcgatcatctgtgaatatgattcaagggttcatactactcttgagacctcaagaccgc	1433
Db	912	aaacatagtcttgaaatgtgagactatgagcaatccacatacttctcagaccctcagaattcc	971
QY	1434	agccggagagacctgtgtctctctcgaaagaacaggaagtgcaatgagctacaccgycggaagaa	1493
Db	972	caccacca--agctccctctgttcgagagaccacacagcagatcaagtttctctacaatgtgca	1030
QY	1494	gagcgctgcagaacagaccctctgcctctgcagcagccctcccgcgagttctggtctccggg	1553
Db	1031	gaactcaccaagaacaccccccaagcttltgacccgggcacactgttctgtgccacacttg	1090
QY	1554	cttctccctccgggcgcacaccgctgacaggttgtaacttgagcttggtggcgacgycgctg	1613
Db	1091	catcaacagggggagacacacacgttggttggtatgatatgaacttgccatctgggggacagt	1150
QY	1614	caaggctgggggtgtgacccggggaggggtgtagagagacaggaagatgagactcagcgccg	1673
Db	1151	caccgtggcggtgtgtgagcgaagaatgtgcacgga--aggggagatcttcgctgctggcgacg	1209
QY	1674	aggacgctgtctgggcgttgatcatctctgcaccaagcagtgtctgggcagacactccc	1733
Db	1210	aggaggggggtgtgggcttgtagagcttggtggcttc--gtctcgctcttggtctcttc	1268
QY	1734	gggacccgaacctgtccgcttgagcagagatccccgcgacgctggaatgcgccttgactac	1793
Db	1269	ccccacagctgtgacctgaaaggagacgccccggcag--gtgaggtgtgtctctgactat	1326
QY	1794	gagggggggcaggtgacacctccacaacgcccagagccaggggcccatactctacactta	1853
Db	1327	gaggtgggctgtgtgtaacttaccacagctgtacaccgagag--cccatctaacacttca	1364
QY	1854	ctgagctcttctctccgccaaggtcttccctgtcccttggtggccgctggaac	1902
Db	1385	ctgctctcttcaactag--aaggtcatctcctcttcttggtctctggggc	1431
RESULT 3			
ID	AAI23948/c		
XX	AAI23948 standard; DNA: 520 BP.		
AC	AAI23948:		
XX	12-OCT-2001 (first entry)		
DT			
XX			
DE	Probe #13881 for gene expression analysis in human cervical cell sample		
XX			
KW	Probe: human: microarray; gene expression; cervical epithelial cell;		
XX	cervical cancer; ss.		
OS	Homo sapiens.		
XX			
PN	WO200157278-A2.		
PD	09-AUG-2001.		
XX			
PF	30-JAN-2001; 2001IWO-US00670.		

XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-488901/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human cervical epithelial cells -  
 PS Claim 25; SEQ ID No 13881; 487pp; English.  
 XX  
 CC The present invention relates to human single exon nucleic acid probes  
 CC (SENP). The present sequence is one such probe. The SENPs are derived  
 CC from human HeLa cells. The SENPs can be used to produce a single exon  
 CC microarray, which can be used for measuring human gene expression in a  
 CC sample derived from human cervical epithelial cells. By measuring gene  
 CC expression, the probes are therefore useful in grading and/or staging  
 CC of diseases of the cervix, notably cervical cancer.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 520 BP; 111 A; 158 C; 137 G; 114 T; 0 other;

Query Match 2.9%; Score 76.4; DB 22; Length 520;  
 Best Local Similarity 57.4%; Pred. No. 8.3e-10;  
 Matches 218; Conservative 0; Mismatches 151; Indels 11; Gaps 4;

QY 1441 gacctggtctctcgtgagagcaggaagtcagtgatcacaccgcgcaggaagagcctg 1500  
 DB 413 GAGCTCTTCTCTGACAGACCGGAGAGTGTGAGGGGGCCCTACAGCAGAGAGTG 354

QY 1501 ccagacagccccctgcgtctcgaagcgtcccgccggtctcgcgggtctcc 1560  
 DB 353 CCGTACACCCAGAGAGATTCGACAGCTGTGTCTCTGCGATGGAGAGCTTCC 294

QY 1561 tccgggcgcacgcgcgtgcaggtgacctgagcttgaggcgagcgcggtgcaggtg 1620  
 DB 293 TCAGGGAACAATTAATCTGGAGG-----TGAGAGTGGAAAACTGATGCTGTG 240

QY 1621 ggggtgcccgggggggtgtagagagacagagagatggagactcagcgccgagagcg 1680  
 DB 239 GGGGTCTGACACACAGTGTGAGAGAAAGG-GAGGTCTCTGTGATTCCTCAGAAATG 181

QY 1681 cgtctggccgtgatactcctcgtacaaagagtgctgggcccagacacctcccgagacc 1740  
 DB 180 CTTCTGACCCCTGGAGATGTTTGA--AACCAATACCGGGCCCTGTCTCCCTG--AGA 125

QY 1741 gacctgcccgtgagcagatcccgccgagcgtagagatgcccctgacctagcagagcg 1800  
 DB 124 GGATTTCTCCTTTGAGAGAGATCCCTTTGCCGGGTGGCGTCTTCTTGACATATGAAGCTG 65

QY 1801 ggcaggtgacctccacacac 1820  
 DB 64 GAGATGTCTCTTCTACAC 45

RESULT 4  
 AAI49253/C  
 ID AAI49253 standard; DNA; 520 BP.  
 XX  
 AC AAI49253;

DT 17-OCT-2001 (first entry)  
 XX  
 DE Probe #17939 used to measure gene expression in human placenta sample.  
 XX  
 KW Probe: microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200157272-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00663.  
 XX  
 PR 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 DR WPI; 2001-488897/53.  
 XX  
 PT Human genome-derived single exon nucleic acid probes useful for  
 XX analyzing gene expression in human placenta -  
 PS Claim 25; SEQ ID No 17939; 654pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.  
 XX  
 SQ Sequence 520 BP; 111 A; 158 C; 137 G; 114 T; 0 other;

Query Match 2.9%; Score 76.4; DB 22; Length 520;  
 Best Local Similarity 57.4%; Pred. No. 8.3e-10;  
 Matches 218; Conservative 0; Mismatches 151; Indels 11; Gaps 4;

QY 1441 gacctggtctctcgtgagagcaggaagtcagtgatcacaccgcgcaggaagagcctg 1500  
 DB 413 GAGCTCTTCTCTGACAGACCGGAGAGTGTGAGGGGGCCCTACAGCAGAGAGTG 354

QY 1501 ccagacagccccctgcgtctcgaagcgtcccgccggtctcgcgggtctcc 1560  
 DB 353 CCGTACACCCAGAGAGATTCGACAGCTGTGTCTCTGCGATGGAGAGCTTCC 294

QY 1561 tccgggcgcacgcgcgtgcaggtgacctgagcttgaggcgagcgcggtgcaggtg 1620  
 DB 293 TCAGGGAACAATTAATCTGGAGG-----TGAGAGTGGAAAACTGATGCTGTG 240

QY 1621 ggggtgcccgggggggtgtagagagacagagagatggagactcagcgccgagagcg 1680  
 DB 239 GGGGTCTGACACACAGTGTGAGAGAAAGG-GAGGTCTCTGTGATTCCTCAGAAATG 181

QY 1681 cgtctggccgtgatactcctcgtacaaagagtgctgggcccagacacctcccgagacc 1740  
 DB 180 CTTCTGACCCCTGGAGATGTTTGA--AACCAATACCGGGCCCTGTCTCCCTG--AGA 125

QY 1741 gacctgcccgtgagcagatcccgccgagcgtagagatgcccctgacctagcagagcg 1800  
 DB 124 GGATTTCTCCTTTGAGAGAGATCCCTTTGCCGGGTGGCGTCTTCTTGACATATGAAGCTG 65

QY 1801 ggcaggtgacctccacacac 1820  
 DB 64 GAGATGTCTCTTCTACAC 45

Dd	64	GAGATGTCCTCCTTACAAAC	45
		RESULT	5
ID	AAI09542/c		
XX	AAI09542	standard; DNA: 520 BP.	
AC	AAI09542;		
XX	09-OCT-2001	(first entry)	
DT			
XX	Probe #9533 used to measure gene expression in human breast sample.		
DE			
KW	Probe; human; breast disease; breast cancer; development disorder; ss;		
RN	Inflammatory disease; proliferative breast disease; non-carcinoma tumour.		
OS	Homo sapiens.		
PX	WO200157270-A2.		
PD			
XX	09-AUG-2001.		
PF			
PR	29-JAN-2001; 2001WO-US00661.		
XX			
PR	04-FEB-2000; 2000US-0180312.		
PR	26-MAY-2000; 2000US-0207456.		
PR	30-JUN-2000; 2000US-0608408.		
PR	03-AUG-2000; 2000US-0632366.		
PR	21-SEP-2000; 2000US-0234687.		
PR	27-SEP-2000; 2000US-0236359.		
PR	04-OCT-2000; 2000GB-0024265.		
XX			
PA	(MOLE-) MOLECULAR DYNAMICS INC.		
XX			
PI	Penn SG, Hanzel DK, Chen W, Rank DR;		
XX			
DR	WPI: 2001-476286/51.		
PT			
PP	Novel single exon nucleic acid probe used to measuring gene expression in a human breast -		
PS	Claim 25; SEQ ID NO 9533; 322pp; English.		
XX			
CC	The present invention relates to novel single exon nucleic acid probes. The present sequence is one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridises at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosting diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer, disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours.		
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences.		
CC			
CC			
SQ	Sequence 520 BP; 111 A; 158 C; 137 G; 114 T; 0 other;		
Query Match	2.9%; Score 76.4; DB 22; Length 520; Best Local Similarity 57.4%; Pred. No. 8.3e-10; Matches 218; Conservative 0; Mismatches 151; Indels 11; Gaps 4;		
OY	1441 gaccggtcttcctcgaagacagcgagtgagtgcattacccggcagagaagacctg	1500	
Dd	413 GAGCTCTTCCTGTCCAGAGACC GGGAAGTGAGGGCGGCCCTTACAGCAAGAGTg	354	
OY	1501 ccagacagcccccttgcgcttcgacgagcctcccgcgagttctggtcccggctctcc	1560	
Dd	353 CCTGACAACCCAGAGAAGATTGCACAGTcAGCCTTGTGCTCTGGGATGGAGAAGCTTCGC	294	
OY	1561 tcgggagcgcaacgcgtgccaagtttaacttgcagcttggcgcaaggcgcggtctgcaggtg	1620	

[illegible]

CC HFE gene sequences from the present invention can be used to develop  
CC products for use in the diagnosis and treatment of HFE. The present  
CC invention also describes BTF genes, which are homologues of the milk  
CC protein butyrophilin (BT), and can be used in the production of agonists  
CC and antagonists of BT function. The present sequence encodes BTF2. Also  
CC described are: (1) a Robket gene which can be used to develop products  
CC for the study, diagnosis and treatment of lupus and Sjogren's syndrome;  
CC and (2) NPT3 and NPT4 genes which are homologues of a type 1 sodium  
CC transport gene, and can similarly be used for hyphophosphatemia.  
XX  
XX Sequence 3518 BP; 926 A; 831 C; 868 G; 893 T; 0 other;

Query Match	2.9%	Score 76.4;	DB 19;	Length 3518;
Best Local Similarity	57.4%;	Pred. No. 2e+09;		
Matches 218; Conservative	0;	Mismatches 151;	Indels 11;	Gaps 4;

[illegible]

RESULT	7
AAC77707	
ID	AAC77707 standard; cDNA; 2981 BP.

AC	AAC77707;
XX	
DT	08-FEB-2001 (first entry)

DE Human cancer associated gene sequence SEQ ID NO:101.

KM Human, cancer associated gene; cancer antigen; detection; cancer;  
KM diagnosis; cytostatic; proliferative; vulnery; immunomodulator;  
KM antidiabetic; antilastmatic; antirheumatic; antiarthritic; antiviral;  
KM antinflammatory; antithyroid; antiallergic; antibacterial; cardiant;  
KM dermatological; neuroprotective; thrombolytic; coagulant; nootropic;  
KM vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation  
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;  
KM allergic reaction; graft versus host disease; organ rejection;  
KM haemostatic; thrombolytic; cardiovascular disorder; infection;  
KM neurological disease; drug screening; ss.

OS	Homo sapiens.
XX	
PN	W0200055350-A1
XX	
PD	21-SEP-2000.
XX	

PF 08-MAR-2000; 2000WO-US05882.

AA 12-MAR-1999; 99US-0124270.  
PR

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

DR WPI; 2000-587533/55.

XX F 5000, 74404200  
XX

PT useful for treating or diagnosing e.g. cancer -

PS Claim 1; Page 691-692; 2352pp; English.

CC AAC776039 to AAC78448 encode the human cancer-associated proteins given  
CC in ABA43398 to ABA44239. The proteins can have activities based on the  
CC tissues and cells the genes are expressed in. Example of activities  
CC include: cytoskeletal; proliferative; vulnerrary; immunomodulator;  
CC antidiabetic; antitastmatic; antihemmatic; antiaathritic;  
CC antidiarramatory; antithyroid; antiallergic; antibacterial; antiviral;  
CC dermatological; neuropretic; cardant; thrombolytic; coaguant;  
CC neuroptic; vasotropic; antipsoriatic and antiangiogenic. The  
CC polypeptides and polypeptides can be used for preventing, treating or  
CC ameliorating medical conditions and diagnosing pathological conditions.  
CC The polypeptides, polypeptides, antibodies, agonists and antagonists from  
CC the present invention may be used to treat immune disorders by activating  
CC or inhibiting the proliferation, differentiation or mobilisation of  
CC immune cells, to treat disorders of haematopoietic cells, autoimmune  
CC disorders, allergic reactions, graft versus host disease and organ  
CC rejection, modulate haemostatic or thrombolytic activity, modulate  
CC inflammation, cancers, cardiovascular disorders, neurological disease and  
CC bacterial or viral infections. The peptides, nucleotides, antibodies,  
CC agonists and antagonists may be also be used in drug screens. AAC78449 to  
CC AAC78457 and ABA44240 represent sequences used in the exemplification of  
CC the present invention.

Sequence 2981 BP; 669 A; 807 C; 796 G; 706 T; 3 other;

Query Match	2.7%	Score 69.2	DB 21	Length 2981
Best Local Similarly	52.6%	Pred. No. 1.3e-07		
Matches 224	Conservative	0	Mismatches 193	Indels 9
				Gaps 3

[illegible]



```

QY 1816 acaacy 1821
Db 1687 acacy 1692

RESULT 8
AA60263
ID AAX60263 standard; DNA; 43226 BP.
XX
AC AAX60263;
XX
DT 12-AUG-1999 (first entry)
XX
DE Nucleic acid sequence from C121 gene of system B of chicken MHC.
XX
KW Resistance; tumour development; Marek disease tumour; chicken;
KW system B; system Rfp-Y; poultry major histocompatibility complex; MHC;
KW class II B-L gene; gene 17.5; gene 12.3; gene B-FIV; class I;
KW genotype; selection; breeding; virus-induced tumour; C121 gene; ss.
XX
OS Gallus sp.
XX
PN FR2771422-A1.
XX
PD 28-MAY-1999.
XX
PF 21-NOV-1997; 97FR-0014669.
XX
PR 21-NOV-1997; 97FR-0014669.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
DR WPI; 1999-329699/28.

PT Chicken nucleic acid involved in controlling tumour susceptibility
PS
XX Claim 2; Fig 5; 49pp; French.
XX
CC The specification describes nucleic acid sequences from genes (or
CC related regions) that encode proteins involved in controlling
CC resistance or susceptibility to development of tumours (e.g. Marek
CC disease tumours) in chickens. The nucleic acid sequences include
CC sequences from gene of systems B or Rfp-Y of the poultry major
CC histocompatibility complex (MHC), other than genes of class II B-L
CC and genes 17.5, 12.3 or B-FIV of class I. The nucleic acid sequences
CC are used to genotype poultry, particularly to select (for breeding)
CC birds resistant to virus-induced tumours. The present sequence
CC represents a nucleic acid sequence from a gene of system B.
XX
SO Sequence 43226 BP; 10365 A; 10522 C; 11244 G; 10981 T; 114 other;

Query Match 2.6%; Score 67.6; DB 20; Length 43226;
Best Local Similarity 55.4%; Pred. No. 1.2e-06;
Matches 237; Conservative 0; Mismatches 179; Indels 12; Gaps 5;

QY 1411 tcaactggaccctccagaccgacgagactggtctctcgtgaagaagaagtc 1470
Db 6583 tgactctggaccagagacgagccacctcg-cctcgctctctccaagaagacagagc 6641
QY 1471 gtggagtcacccgcaagaagagccttcgacagacgcccctcgcttcgacgagc 1530
Db 6642 gtccgagtgagatacagcttcgagaaatccccgacgagcccgagcgttcgacgcgat 6701
QY 1531 ccggcggttcggcttcgggtctctcctcgggacgacgacgctggcaggttgaacctg 1590
Db 6702 cccgcgtgctgggttggaaccttcacctcctcggaagcactgctgggtg-----tg 6755
QY 1591 cagctggcgacgagcgagcggtgacaggtgggtggccggggaggggtgaagagagca 1650
Db 6756 gatctacagaagagcgactgacgctgggggtgacagagaggtccctgcacagga-a 6814
QY 1651 gggagagatggactcagcgcgagagcgagcgtctggccgtgatcatctctgcaccaa 1710

```

```

Db 6815 aggagcgctcagcttaacccttgatgaagcactctgggctgty--cagcaatgggggttc 6872
QY 1711 cagtgctgggacgacactcccggggacgacgactgtctcgtgagagatccgcgcag 1770
Db 6873 aagacagagcctcactccctcccgaccacactgaacactccacgggttcccaaaag 6932
QY 1771 gcgtgagagtcgacctgactacgagcgaggcgaggtgacctccaacagcccagagcc 1830
Db 6933 --atcgcactctctcgactacgaatggcgagtggtgcttlttgatgtgagaacc 6990
QY 1831 caggggcc 1838
Db 6991 aatggccc 6998

RESULT 9
AA78067
ID AAX78067 standard; cDNA; 712 BP.
XX
AC AAX78067;
XX
DT 14-NOV-2000 (first entry)
XX
DE cDNA encoding human colon tumour polypeptide, SEQ ID NO:354.
XX
KW Human colon tumour polypeptide; tumour antigen; cancer; vaccine;
KW immunotherapy; diagnosis; progression; ss.
XX
OS Homo sapiens.
XX
PN WO200037643-A2.
XX
PD 29-JUN-2000.
XX
PF 23-DEC-1999; 99WO-US30909.
XX
PR 23-DEC-1998; 98US-0221298.
PR 02-JUL-1999; 99US-0347496.
PR 22-SEP-1999; 99US-0401064.
PR 19-NOV-1999; 99US-0444242.
PR 02-DEC-1999; 99US-0454150.
XX
PA (CORI-) CORIAX CORP.
XX
PI Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolk J;
PI Wang T, Yugu J;
XX
DR WPI; 2000-442671/38.
XX
PT New colon tumor polypeptides used to inhibit the development of cancer,
PT especially colon cancer, and for diagnosing and monitoring the
PT progression of the cancer -
XX
PS Claim 29; Page 193-194; 229pp; English.
XX
CC Sequences AAX77722-A78199 represent 478 cDNAs encoding proteins or
CC portions of proteins which are associated with human colon tumours.
CC The invention also specifically discloses 8 human colon tumour proteins
CC (AAB11897-B11904). The nucleic acids, the polypeptides they encode, and
CC antigen presenting cells (APCs, preferably dendritic cells) expressing
CC such polypeptides may be used in vaccines that target tumour cells,
CC especially colon tumour cells, thereby inhibiting the development of
CC cancer. T-cells specific for the polypeptide expressed by the APC are
CC used to remove tumour cells from biological samples, especially blood or
CC fractions thereof. The sample or the isolated T-cells specific for the
CC polypeptide can then be used to inhibit cancer development. CD4+ and/or
CC CD8+ T-cells from a patient may be incubated with a polypeptide or
CC nucleic acid of the invention, or an APC expressing such a polypeptide,
CC to cause the proliferation of specific T-cells. The T-cells can be
CC cloned and then administered back to the patient to inhibit cancer
CC development. Nucleic acids encoding the polypeptides and antibodies
CC against the polypeptides may be used to determine the expression level

```



PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX WPI; 1999-329699/28.  
 DR Chicken nucleic acid involved in controlling tumour susceptibility  
 XX  
 PT  
 XX  
 PS Claim 2; Fig 5; 49pp; French.  
 XX  
 CC The specification describes nucleic acid sequences from genes (or  
 CC related regions) that encode proteins involved in controlling  
 CC resistance or susceptibility to development of tumours (e.g. Marek  
 CC disease tumours) in chickens. The nucleic acid sequences include  
 CC sequences from gene of systems B or Rfp-Y of the poultry major  
 CC histocompatibility complex (MHC), other than genes of class II B-L  
 CC and genes 17.5, 12.3 or B-FIV of class I. The nucleic acid sequences  
 CC are used to genotype poultry, particularly to select (for breeding)  
 CC birds resistant to virus-induced tumours. The present sequence  
 CC represents a nucleic acid sequence from a gene of system B.  
 CC  
 XX  
 SQ Sequence 43226 BP; 10365 A; 10522 C; 11244 G; 10981 T; 114 other;  
 30  
 Query Match 2.48; Score 63; DB 20; Length 43226;  
 Best Local Similarity 53.8%; Pred. No. 1.7e-05;  
 Matches 220; Conservative 0; Mismatches 180; Indels 9; Gaps 4;  
 OY 1402 caggagtcatactactctgaccctcagaccgacgagacgctgtctctctggaagac 1461  
 Db 36012 CAGCCAGCATACCTCGACCCCGACCCGCTCACCC-CTGACCTCATCTCTCCGAAAGAC 35954  
 OY 1462 aggaagtcagtgaggtacacccgagcagaagaagagcctgcacagaccctgcgcttc 1521  
 Db 35953 CGAAGAAGTGTGAACGTTGGGGAAGACACAGACGACTTCTGATTAACCCGGAAGATTC 35894  
 OY 1522 gacgagctcccgagcggtctgtggtctcccgagctctctcccgagcgccacgctgcag 1581  
 Db 35893 GCCTACTGCGCCCTTGTGTTTGGGCCACCAAGCTTCTGCGCGCGCCTGCTGGGAG 35834  
 OY 1582 gttagcctgacgctgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1641  
 Db 35833 G-----TGGAAAGTGGGGATGAAGAGGACTGGCCATCGGTGTGGCCGGAATCCATC 35780  
 OY 1642 aggaagcagaagagatgagactcagcgacgagcgagcgagcgagcgagcgagcgagc 1701  
 Db 35779 CCTGGGA-AGGGTCAACTACGCTCTGTCCCAAGGGGGGATTGGGGGGGAGGAATG. 35721  
 OY 1702 tgcaccaagcagtgctgagcgacgacgtcccgagcgacgagcgagcgagcgagcgag 1761  
 Db 35720 GGGGGGACAGGTCCGGGCACTCACCCACAGGTGACCTGTAGCTTGCCTGGGT 35661  
 OY 1762 ccgagcgagcgagtgagagtgccctgagctacgagcgagcgagcgagcgagcgagc 1810  
 Db 35660 GCCCAGGAGG-GTCAAGCTCCACTTGGACTGAGGTGAGGAGCGGTGC 35613  
 RESULT 12  
 AAH72187  
 ID AAH72187 standard; CDNA: 376 BP.  
 XX AAH72187;  
 AC  
 XX 19-SEP-2001 (first entry)  
 DT  
 XX  
 DE Human cervical cancer marker nucleic acid 3461.  
 XX  
 KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200142467-A2.  
 XX  
 PD 14-JUN-2001.  
 XX

PF 08-DEC-2000; 2000WO-US33312.  
 XX  
 PR 08-DEC-1999; 99US-0169681.  
 PR 21-DEC-1999; 99US-0171350.  
 PR 14-MAR-2000; 2000US-0189315.  
 PR 12-MAY-2000; 2000US-0203791.  
 PR 09-JUN-2000; 2000US-0210600.  
 PR 21-JUL-2000; 2000US-0220114.  
 XX  
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
 XX  
 PI Schlegel R, Deeds J, Berger A, Zhao X;  
 PI  
 DR WPI; 2001-375006/39.  
 XX  
 PT New isolated nucleic acid for diagnosing and treating cervical cancer  
 PT and for assessing and detecting compounds for treating the cancer -  
 XX  
 PS Claim 1; Page 663-664; 1051pp; English.  
 XX  
 CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
 CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
 CC polypeptides are useful: to assess if a patient is afflicted with  
 CC cervical cancer or has a pre-malignant condition; to monitor the  
 CC progression of cervical cancer or a premalignant condition in a patient;  
 CC to select and/or assess the efficacy of a compound or therapy for  
 CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
 CC useful for gene therapy.  
 CC  
 XX  
 SQ Sequence 376 BP; 84 A; 94 C; 107 G; 91 T; 0 other;  
 50  
 Query Match 2.38; Score 60.6; DB 22; Length 376;  
 Best Local Similarity 60.78%; Pred. No. 7.9e-06;  
 Matches 116; Conservative 0; Mismatches 74; Indels 1; Gaps 1;  
 OY 1400 ttcaagggtcactcctcctgacccctcagaccgacgagcgagcgagcgagcgagcgag 1459  
 Db 64 tacagtgagctgactctgagaccgacgagcgagcgagcgagcgagcgagcgagcgag 122  
 OY 1460 acaggaagtcagtgaggtacacccgagcagaagaagagcctgcacagacgagcgagcg 1519  
 Db 123 atctcgagcaagtgaggtacaggtaccccaacagagcctgctacacacccgagaggt 182  
 OY 1520 tcgagcgctcccgagcggttcttgagcttcccgagcttctctcccgagcgagcgagcg 1579  
 Db 183 tcaatctgttccctgtgtcttggtctctcatgcttcatcgcgagcgagcgagcgagcg 242  
 OY 1580 aggtgacctg 1590  
 Db 243 aggtagaggtg 253  
 RESULT 13  
 AAH71127  
 ID AAH71127 standard; CDNA: 509 BP.  
 XX AAH71127;  
 AC  
 XX 19-SEP-2001 (first entry)  
 DT  
 XX  
 DE Human cervical cancer marker nucleic acid 2401.  
 XX  
 KM Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200142467-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 08-DEC-2000; 2000WO-US33312.  
 XX

```
PR 08-DEC-1999: 99US-0169681.
PR 21-DEC-1999: 99US-0171350.
PR 14-MAR-2000: 2000US-0189315.
PR 12-MAY-2000: 2000US-0203791.
PR 09-JUN-2000: 2000US-0210600.
PR 21-JUL-2000: 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI: 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1, Page 504; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 509 BP; 103 A; 131 C; 149 G; 126 T; 0 other;

Query Match      2.3%; Score 60.6; DB 22; Length 509;
Best Local Similarity 60.7%; Pred. No. 9.1e-06;
Matches 116; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 1400 ttgaagggtacatcactctgacccctcagaccgcagccgagacccgtgtctctcggaag 1459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 94 taacagtgagcagtgactctctgacccagacagcagcgtactccag-cctgactctctcgata 152

QY 1460 acaggaagtcagtgaaggtacacccgcgaagaagaagcctgcagacagccctcgct 1519
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 153 atctgcgcagcagtgctgctacagttaccctcaacagacgacctgctgacaaaccgagaggt 212

QY 1520 tcgacggccctccggcggtcttcggcttcggcgtctctcctcggcgccacgcgcgcgc 1579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 213 tcaatctgttccctgtctcttggtctctcatgtctcattcgcgcggagacattatggg 272

QY 1580 aggttgacctg 1590
   ||| ||| |||
DB 273 aggtagaggtg 283

RESULT 14
AAH72832
ID AAH72832 standard; cDNA; 750 BP.
XX
AC AAH72832;
XX
DT 19-SEP-2001 (first entry)
XX
DE Human cervical cancer marker nucleic acid 4106.
XX
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200142467-A2.
PD 14-JUN-2001.
XX
PF 08-DEC-2000; 2000WO-US33312.
XX
PR 08-DEC-1999: 99US-0169681.
PR 21-DEC-1999: 99US-0171350.
```

```
PR 14-MAR-2000; 2000US-0189315.
PR 12-MAY-2000; 2000US-0203791.
PR 09-JUN-2000; 2000US-0210600.
PR 21-JUL-2000; 2000US-0220114.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Deeds J, Berger A, Zhao X;
XX
DR WPI: 2001-375006/39.
XX
PT New isolated nucleic acid for diagnosing and treating cervical cancer
PT and for assessing and detecting compounds for treating the cancer -
XX
PS Claim 1, Page 876-877; 1051pp; English.
XX
CC The invention relates to novel genes (AAH68727-AAH73383) associated with
CC cervical cancer with cytostatic activity. The nucleic acids and encoded
CC polypeptides are useful: to assess if a patient is afflicted with
CC cervical cancer or has a pre-malignant condition; to monitor the
CC progression of cervical cancer or a premalignant condition in a patient;
CC and to select and/or assess the efficacy of a compound or therapy for
CC inhibiting cervical cancer in a patient. The nucleic acids may also be
CC useful for gene therapy.
XX
SQ Sequence 750 BP; 191 A; 167 C; 196 G; 192 T; 4 other;

Query Match      2.3%; Score 60.6; DB 22; Length 750;
Best Local Similarity 60.7%; Pred. No. 1.1e-05;
Matches 116; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 1400 ttgaagggtacatcactctgacccctcagaccgcagccgagacccgtgtctctcggaag 1459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 438 tgaagtgagcagtgactctctgacccagacagcagcgtactccag-cctgactctctcgata 496

QY 1460 acaggaagtcagtgaaggtacacccgcgaagaagaagcctgcagacagccctcgct 1519
   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 497 atctgcgcagcagtgctgctacagttaccctcaacagacgacctgctgacaaaccgagaggt 556

QY 1520 tcgacggccctccggcggtcttcggcttcggcgtctctcctcggcgccacgcgcgcgc 1579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 557 tcaatctgttccctgtctcttggtctctcatgtctcattcgcgcggagacattatggg 616

QY 1580 aggttgacctg 1590
   ||| ||| |||
DB 617 aggtagaggtg 627

RESULT 15
AAV05159
ID AAV05159 standard; DNA; 14180 BP.
XX
AC AAV05159;
XX
DT 20-JUL-1998 (first entry)
XX
DE Mouse butyrophilin gene.
XX
DE Butyrophilin; b1n gene; promoter; mouse; transgenic animal;
KW mammary gland-specific gene expression; breast cancer; diagnosis;
KW ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT Promoter 1..4693
   /*tag= a
FT TATA_signal 4645..4651
   /*tag= b
FT prim_transcript 4611
   /*tag= c
```

```
FT /evidence= EXPERIMENTAL
FT /standard_name= transcription start site
FT /note= "most frequently used site"
FT misc_signal
FT /tag= d
FT /note= "AT-rich, TATAA-like element"
FT misc_signal
FT /tag= e
FT /note= "AT-rich, TATAA-like element"
FT misc_signal
FT /tag= f
FT /note= "CCAAT-like element"
FT misc_signal
FT /tag= g
FT /note= "CCAAT-like element"
FT misc_signal
FT /tag= h
FT /note= "CCAAT-like element"
FT prim_transcript
FT /tag= i
FT /evidence= EXPERIMENTAL
FT /standard_name= transcription start site
FT prim_transcript
FT /tag= j
FT /evidence= EXPERIMENTAL
FT /standard_name= transcription start site
FT polyA_signal
FT /tag= k
FT /evidence= EXPERIMENTAL
FT prim_transcript
FT /tag= l
FT /evidence= EXPERIMENTAL
FT /standard_name= 3' end of transcript
FT CDS
FT /tag= m
FT /evidence= EXPERIMENTAL
FT /note= "contains introns"
FT sig_peptide
FT /tag= n
FT /evidence= EXPERIMENTAL
FT CDS
FT /tag= o
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT CDS
FT /tag= p
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT CDS
FT /tag= q
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT CDS
FT /tag= r
FT /evidence= EXPERIMENTAL
FT /standard_name= translational initiation codon
FT exon
FT /tag= s
FT /codon_start= 4923..4925
FT intron
FT /tag= t
FT /evidence= EXPERIMENTAL
FT /number= A
FT exon
FT /tag= u
FT /number= 2
FT intron
FT /tag= v
FT /evidence= EXPERIMENTAL
FT /number= B
FT exon
FT /tag= w
FT /number= 3
FT intron
FT /tag= x
FT /evidence= EXPERIMENTAL
FT /number= C
FT exon
FT /tag= y
FT /number= 4
FT intron
FT /tag= z
FT /evidence= EXPERIMENTAL
FT /number= D
FT exon
FT /tag= aa
FT /number= 5
FT intron
FT /tag= ab
FT /evidence= EXPERIMENTAL
FT /number= E
FT exon
FT /tag= ac
FT /number= 6
FT intron
FT /tag= ad
FT /evidence= EXPERIMENTAL
FT /number= F
FT exon
FT /tag= ae
FT /number= 7
FT repeat_unit
FT /tag= af
FT /rpl_type= INVERTED
FT repeat_unit
FT /tag= ag
FT /rpl_type= INVERTED
FT protein_bind
FT /tag= ah
FT /note= "NFI"
FT protein_bind
FT /tag= ai
FT /note= "gamma-IRE"
FT protein_bind
FT /tag= aj
FT /note= "myb"
FT protein_bind
FT /tag= ak
FT /note= "YY1"
FT protein_bind
FT /tag= al
FT /note= "gamma-IRE"
FT protein_bind
FT /tag= am
FT /note= "YY1"
FT protein_bind
FT /tag= an
FT /note= "NFI"
FT protein_bind
FT /tag= ao
FT /note= "STAT"
FT protein_bind
FT /tag= ap
FT /note= "GMCSP"
FT protein_bind
FT /tag= aq
FT /note= "C/EBP"
FT protein_bind
FT /tag= ar
FT /note= "C/EBP"
FT protein_bind
FT /tag= as
FT /note= "PU.1"
FT protein_bind
FT /tag= at
FT /note= "alpha-IFN"
```

```
FT protein_bind 3195..3203
FT /tag= au
FT /note= "GMCSF"
FT protein_bind 3253..3259
FT /tag= av
FT /note= "gamma-IRE"
FT protein_bind 3370..3378
FT /tag= aw
FT /note= "GRE"
FT protein_bind complement (3397..3406)
FT /tag= ax
FT /note= "GRE"
FT protein_bind complement (3399..3403)
FT /tag= ay
FT /note= "gamma-IRE"
FT protein_bind 3419..3427
FT /tag= az
FT /note= "STAT"
FT protein_bind 3439..3445
FT /tag= ba
FT /note= "gamma-IRE"
FT protein_bind 3450..3454
FT /tag= bb
FT /note= "GMCSF"
FT protein_bind 3456..3462
FT /tag= bc
FT /note= "STAT"
FT protein_bind 3457..3463
FT /tag= bd
FT /note= "C/EBP"
FT protein_bind complement (3464..3468)
FT /tag= be
```

```
Query Match 2.3%; Score 60.6; DB 19; Length 14180;
Best Local Similarity 54.6%; Pred. No. 4.3e-05;
Matches 230; Conservative 0; Mismatches 179; Indels 12; Gaps 5;

QY 1400 ttcagggtatcaactctgacccctcagaccgcgacgacgtgtctctcgaag 1459
Db 10728 tccagttgacgtgactctgatacagacagccacc-cccactctctctgatatgaag 10786

QY 1460 acaggaagtcagtgaggtacacccgcgagagaagagcctgcccagagccctgcct 1519
Db 10787 attcaagaagtcagtcgacttggaagattcacgctcagatccctgctgatatagacagagaagat 10846

QY 1520 tcgacggcctcccgcggtcttggtctcccggtctctcctccggcgccacgcgtgac 1579
Db 10847 ttgactcttgccctggtgtgttgccgtgagaccttacttcaaggagagataactgtgg 10906

QY 1580 aggttgacctgcagctgcgcagcgcgcgctgcacggtcggggtggtggcggtggggg 1639
Db 10907 aggttgg-----aggtggagatatagactgactggccattggtgtgtagggagaatg 10960

QY 1640 tggaggagagagagagatggagctcagcgccgagagcggtctgtggcgctgatatc 1699
Db 10961 tggttgagaagaagggtttgaccca-tgactccgataatagggttcctgggtctgtgaggtt 11019

QY 1700 tcgtcaccaagcagtgctggcgccagcactcccgggcacgacgtgtccgtgagcgag 1759
Db 11020 tatgga--aatgggtactggtgcccctacccactcaaggacc--tctctcgattagcagg 11075

QY 1760 atcccgacgagcgctgagagtcgccttgactcagagcggggcaagtgacccctcaca 1819
Db 11076 gccccctcgagagtggtgtttctctgactatgacgagagacattccttctacaa 11135

QY 1820 c 1820
Db 11136 c 11136
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1765.4	67.6	2255	9	HSJ34249	U34249 Human putat
2	1301.4	49.9	1734	9	AF220132	AF220132 Homo sapi
3	971.2	37.2	47777	9	AF220132	AF220132 Homo sapi
4	971.2	37.2	100000	9	AC004189	AC004189 Homo sapi
5	971.2	37.2	162383	9	AP000517	AP000517 Homo sapi
6	971.2	37.2	162383	9	AB023054	AB023054 Homo sapi
7	714.4	27.4	17284	9	AB023055	AB023055 Homo sapi
8	332.4	12.7	152211	4	HSR1NG6	Y07829 Homo sapien
9	271.4	10.4	318	9	SC551829	ALJ51829 Sus scrof
10	246.8	9.5	426	9	HS73B8R	ALJ51829 Sus scrof
11	198.8	7.6	158414	10	AF220133	AF220133 Homo sapi
12	191.6	7.3	2002	9	AC005603	AC005603 Homo musci
13	170.4	6.5	2269	10	AF220121	AF220121 Homo sapi
14	165.6	6.3	2214	10	AF134811	AF220121 Mus musci
15	160.6	6.2	54674	2	RM499P20	AF134811 Mus musci
16	137.2	5.3	3031	9	AF220123	AL603720 Rattus no
17	117.6	4.5	2151	9	AF396651	AF220123 Homo sapi
18	116	4.4	3595	9	HSU09825	AF396651 Homo sapi
19	105.8	4.1	47777	9	AC004189	Y08825 Human acid
20	105.8	4.1	100000	9	AP000517	AC004189 Homo sapi
21	105.8	4.1	152211	4	SC551829	AP000517 Homo sapi
22	105.8	4.1	162383	9	AB023054	ALJ51829 Sus scrof
23	105.8	4.1	162383	9	AB023055	AB023054 Homo sapi
24	102.6	3.9	17284	9	HSR1NG6	AB023055 Homo sapi
25	98	3.8	281	9	HS73B8R	Y07829 Homo sapien
26	97.4	3.7	1888	9	AF396652	Z63789 H.sapiens C
27	97.4	3.7	1949	9	AF396653	AF396652 Homo sapi
28	97.4	3.7	1956	9	AF396653	AF396653 Homo sapi
29	97.4	3.7	2338	9	AF396655	AF396653 Homo sapi
30	92.8	3.6	789	9	HSU00221	AF396655 Homo sapi
31	90.2	3.5	120524	9	AC008443	AL360165 Homo sapi
32	89.6	3.4	236933	2	AL161615	AC008443 Homo sapi
33	84.4	3.2	2256	9	HSU000220	AL161615 Homo sapi
34	82.8	3.2	174612	2	AC023889	AL360163 Homo sapi
35	81.4	3.1	1821	9	BC011629	AC023889 Homo sapi
36	79.4	3.0	183399	2	AC008620	BC011629 Homo sapi
37	76.6	2.9	158414	10	AC005603	AC008620 Homo sapi
38	76.4	2.9	3502	6	AR036568	AC005603 Homo musci
39	76.4	2.9	3578	6	HSU090550	AR036568 Sequence
40	76.4	2.9	39412	6	AY5300	U90500 Human butyr
41	76.4	2.9	170001	6	AY5300	AY5300 Sequence 31
42	76.2	2.9	134331	2	RM462P21	AL021917 Human DNA
43	75	2.9	164766	2	AC026657	AL603727 Rattus no
44	75.4	2.9	54674	2	RM499P20	AC026657 Homo sapi
45	69.2	2.7	1782	9	HU08FPA	AL603720 Rattus no
						J03407 Human r1P t

## ALIGNMENTS

RESULT	1
LOCUS	HSUJ34249
DEFINITION	Human putative zinc finger protein (ZNF7) mRNA, complete cds.
ACCESSION	U34249
VERSION	U34249.1 GI:4096653
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	Geel,V.L., Parmoo,S., Caposela,A., Chu,T.W. and Gruen,J.R.
TITLE	Isolation of novel non-HA gene fragments from the hemochromatosis region (6p21.3) by cDNA hybridization selection
JOURNAL	Am. J. Hum. Genet. 54 (2), 244-251 (1994)
MEDLINE	94136483
REFERENCE	2 (bases 1 to 2255)
AUTHORS	Geel,V.L., Kuida,S., Chu,T.W., Bowlus,C. and Gruen,J.R.
TITLE	A cluster of genes proximal to HLA-A encode putative zinc finger

JOURNAL proteins  
 REFERENCE unpublished (1995)  
 AUTHORS 3 (bases 1 to 2255)  
 TITLE Goel, V. L.  
 JOURNAL Direct Submission  
 FEATURES Submitted (17-AUG-1995) Vita Goel, Pediatrics, Yale University  
 School of Medicine 333 Cedar Street, New Haven, CT 06510-8064, USA  
 Location/Qualifiers

3'UTR	1892.	.2255

Query Match	67.68	Score 1765.4	DB 9	Length 2255
Best Local Similarity	93.08	Pred. No. 0		
Matches 2142	Conservative	0	Mismatches 111	Indels 50
				Gaps 26

QY	61	gctctgtccccaattctccccaacacatggtggtctaaagttctcaagaaaccaatccc	120
Db	61	gcttctgctcccccaattctccccaacacatggtggtctaaagttctcaagaaaccaatccc	120
QY	121	cctgtgaccccaagggtgtgtctctgtctctccacctagaaggacagagaagccaggatlaa	180
Db	121	cctgtgaccccaagggtgtgtctctgtctctccacctagaaggacagagaagccaggatlaa	180
QY	181	agaactgtgcttaacttggccgcgcacatcgtggaaattctgtgttaattcagacgccttggaatt	240
Db	181	agaactgtgcttaacttggccgcgcacatcgtggaaattctgtgttaattcagacgccttggaatt	240
QY	241	tggaccacactcgcgtgatagtggtgtgggcagggttcttagggaaacaaagaagccgaagcca	300
Db	241	tggaccacactcgcgtgatagtggtgtgggcagggttcttagggaaacaaagaagccgaagcca	300

Oy	301	ggtggtctccctgtgctggaatctctgtgctctctctctctctctctctct	360
Db	301	ggtggtctccctgtgctggaatctctgtgctctctctctctctctctctct	360
Oy	361	ctctctctctctctgctctctcagaactctgaaagccgctttccctccctgctgctcaatgtg	420
Db	361	ctctctctctctctgctctctcagaactctgaaagccgctttccctccctgctgctcaatgtg	420
Oy	421	agcttgactcgatctcaaggaaaggaaactcgcgtgctgtgagggagaccggagatggaag	480
Db	421	agcttgactcgatctcaaggaaaggaaactcgcgtgctgtgagggagaccggagatggaag	480
Oy	481	gctgggaagagcaacgctgaaatgccccgaacccccctccctggaagggatgctcaatgac	540
Db	481	gctgggaagagcaacgctgaaatgccccgaacccccctccctggaagggatgctcaatgac	540
Oy	476	gctgggaagagcaacgctgaaatgccccgaacccccctccctggaagggatgctcaatgac	530
Db	476	gctgggaagagcaacgctgaaatgccccgaacccccctccctggaagggatgctcaatgac	530
Oy	541	tgcctgcccctgtaaccctctgtgctggggccgctgtgaggaatgctgtaaccatccctgtgac	600
Db	541	tgcctgcccctgtaaccctctgtgctggggccgctgtgaggaatgctgtaaccatccctgtgac	600
Oy	601	aacacctctgcccgcctctgtctcccccgcgctctccagaatggggagccaatctctgtgac	660
Db	601	aacacctctgcccgcctctgtctcccccgcgctctccagaatggggagccaatctctgtgac	660
Oy	591	ACACTTCTTGCCTGCTGCTGCTCCCGCTCTCCAGATGGGGGCCAATCTTG -GGC	649
Db	591	ACACTTCTTGCCTGCTGCTGCTCCCGCTCTCCAGATGGGGGCCAATCTTG -GGC	649
Oy	661	aagaatctgctctgcccgcctctgccaagagatgaagcagcagaactccatgcccct	720
Db	650	AAGATTCCTGCTCTCCGCTCTCCAGAGAGAGACAGCAGACATCCATGGCCCT	709
Oy	721	gtgccccttgagcccgctgggaagataactactactgctgagggagccggcggaagatctact	780
Db	710	GTGCCCTCGGGCCCGCTGGGAGAG -AATCTTACTGCGAGGAGCAGCGCGAAGATCTACTT	768
Oy	781	ctctctgcaagaagatgacggaatctctctctgtgtgtctgcaagagaggtcccaacgacac	840
Db	769	CTTC -TGCAGAAAGATGCGGAGTTCCTCTGTGTGTCGAGGGAGGCTCCACAGCAC	827
Oy	841	agggcgcaacccgtgggggtctccctggaagggacatcaagccctaacgggagatgctctcaga	900
Db	828	AGGGCGCACCCGTGGGGTCTCTGAGAGAGCCATTCAGCCCTACCGGATCTCTTCACAGA	887
Oy	901	gtcgaactggaagctctgaagcagcagagagaatgagaatgtgaatgtcaaatggtctcaaga	960
Db	888	GTGCACGTGGAAGCTCTAGAGCAGGAGAGATGAGATTG -AAGATGTAAAGTGTCAAGA	946
Oy	961	gaaccaagaactcaaatgtcgctgtgaactcagatctgaacaagacaagaagccgtcaaggtgtca	1020
Db	947	GACCAGAGACTTCAAAGTGCCTGACTCAGATTCGA -AAGCAAGAAGCTCAGGTGGGAA	1004
Oy	1021	cacagctcccttgaaagagctgtcaagcggagagctgcaagcaacagatgtctccctgtgtgcg	1080
Db	1005	CA---GCTTTTGAAGAGCTGC -AGCAGAGAGCTGAGACACAGCATGTCTCTGCTGGCC	1066
Oy	1081	caagaactgaaatgtaacgctctcgagagcaagaatttgaaagagaggaatgaatatataca	1140
Db	1061	AGCGTTGAGGGAGCT-----GAGACAGCAGATTTTGAAGAGAGGATGAATATATACAA	1114
Oy	1141	aaggtctctgaagaaagtcacccggctcttgagccccagcctcaagagagctcgaagaggaagt	1200
Db	1115	AAGGTCTCTGAGGAAGTCAACCCGGCTTTGAG -CCCAAGTCTTAAGGAGCT -GGAGGGAAGT	1172
Oy	1201	gtcagcagcagcaagatgagactctcaagaatgtcgaagtcaagcagaagcagagctgtag	1260
Db	1173	GTCAGCAGCCAGCAAGTGAAGCTTCTAAGATGTCAAGATCA -CCAAGCAGAGTGTAG	1231
Oy	1261	atgaagaacttgtgagctctgagagccaatctctccctgacctgttcaagaagaatccgtga	1320
Db	1232	ATGGAAGACTTTTGTGATCTCTGAGGCCATTTCT -CCTGACCTTTGTCAAGAGATCCGTGA	1290
Oy	1321	tttccagaagaaaaatactaacctctccagaagatgagaaatggtcttcaagaagaactctgg	1380
Db	1291	TTTTCCACAGAAATACTCACTCTCCACAGATGATGAGAGATGTTCT -AGAAATCTTG	1349



QY 1381 cgcatactgtgaaatagattcaagggtgcatcactgtgacctcagaccgacccgacgga 1440  
 |||||  
 Db 1380 CGCATCATCTGGAATATATTTCAGGGGTATCATCTGTGACCTCTAGACCCGACCCGGA 1409  
 QY 1441 gaactgtctctcgtggaagacaggaagtcagtgagtaacccggcagaagaagacctg 1500  
 |||||  
 Db 1440 G-CTGTTCTCTGCGGAGAGACAGAAAGTCAAGAGTACACCCGCGAGAGAGAGCTG 1468  
 QY 1501 ccaaacagagccccctgggtctgtagaggtcccggtctccgggtctccgggtctcc 1560  
 |||||  
 Db 1469 CCAAGACAGCCCCCTGCTGAGCGGCTCCCGGGGTTCTGCGGCTCCCGGCTTCTCC 1528  
 QY 1561 tcgagcgccacacgctgtagaggtttagctgtagctgtagcgagcgagcgagctgtag 1620  
 |||||  
 Db 1529 TCCGGGCCCCACCGTGGCAGGTTACCTGCAGCTGGGCGAGCGGGGGCTGCACGGTG 1588  
 QY 1621 ggggtgagcgagggggtgaggaagagagagagagagagagagagagagagagag 1680  
 |||||  
 Db 1589 GGGGTGGCGGGGAGGCGGTGAGGAGGA-AGGAGAGATGGGACTCAGCGCGAGAGACGG 1647  
 QY 1681 cgtctggagcgtgtatcatctctgtagcacaagagagagagagagagagagagag 1740  
 |||||  
 Db 1648 GGTGAGGCGCTGATCATCTC--GCACCAAGAGTGTGGCCAGACCTCCCGGGGACG 1705  
 QY 1741 gacctgtcagctgtagagagatcccgagcgagagagagagagagagagagagagag 1800  
 |||||  
 Db 1706 GACCTG-CCGCTGAGCGAGATCCCGCG-GGCGTAGAGTGGCCTTGAGATACGAGGCGG 1763  
 QY 1801 ggcaggtgacctccacacagccagagagagagagagagagagagagagagagagag 1860  
 |||||  
 Db 1764 GCGAGGTACCCCTCCACAAACCCCGA--CCAGAGAGCCATCTTACACTTACAGCTCC 1821  
 QY 1861 tttctccgagcaaggtcttccctgtctctgtgagcgagagagagagagagagagag 1920  
 |||||  
 Db 1822 TTT--CTCCGGCAAGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1876  
 QY 1921 agctgtacagcgaggaatgagggcgagagagagagagagagagagagagagagag 1980  
 |||||  
 Db 1877 AGCTGA-AAAGGTGAAGTGGGGCGCGGAGAGGCGGAGAGAGAGAGAGAGAGAGAG 1935  
 QY 1981 ggcagtcagctccgccccctgagcagtgtagcgccccgggagctccctgtgccccg 2040  
 |||||  
 Db 1936 GGGATCCAGCTCCGCCCTGCGCACTGTGCGGCCGGGGCTCCCTGCGCCGCTGAGG 1995  
 QY 2041 cgaagagaaacgggagctgagctgtagcgaagagagagagagagagagagagag 2100  
 |||||  
 Db 1996 CGAAGAA--CAGGGGACTTGAAGTCTGAAACAGCGGTTGTTTATCTTATCTTAG 2053  
 QY 2101 ggcctcagctccctgagctgctgtagcctccctgtgagcctgtagcctctctgaccca 2160  
 |||||  
 Db 2054 GCCCTCAGCTCCCTGACCTCTGAGCCCTCCTGTGACGCTGTGCGCTTCTGTGCACTCA 2113  
 QY 2161 gagtgcagaacacagagagctgagctgtgtaggcaaacagcaactagagagagag 2220  
 |||||  
 Db 2114 GAGTGCAGAAACACAGAGAGGCTTGGCTGTGCTGAGGCAACAGCAACTAGAGAGCAG 2173  
 QY 2221 ccggccttgggggaaaaaaccaagaagagagagagagagagagagagagagagag 2280  
 |||||  
 Db 2174 CGGGCTTTCGGGGAAAAA---AAAGAAAAAGACATTAATAATGTTTAAACTGTTT 2230  
 QY 2281 caagataatlatcttgggaaaaa 2303  
 |||||  
 Db 2231 CAAATATAAAAAAAAAAAAAA 2253

RESULT 2  
 AF220132 1734 bp mRNA PRI 18-MAY-2001  
 LOCUS Homo sapiens tripartite motif protein TRIM15 alpha (TRIM15) mRNA,  
 DEFINITION complete cds; alternatively spliced.  
 ACCESSION AF220132  
 VERSION AF220132.1 GI:12407432  
 KEYWORDS

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eultheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1734)  
 AUTHORS Reymond,A., Meroni,G., Fantozzi,A., Merla,G., Calro,S., Luzzi,L.,  
 Riganelli,D., Zanaria,E., Messali,S., Calmarca,S., Guffanti,A.,  
 Minucci,S., Pelicci,P.G. and Ballabio,A.  
 TITLE The tripartite motif family identifies cell compartments  
 JOURNAL EMBO J. 20 (9), 2140-2151 (2001)  
 REFERENCE 2 (bases 1 to 1734)  
 AUTHORS Reymond,A. and Meroni,G.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-DEC-1999) TIGEM, Via Olgettina 58, Milan, MI 20132,  
 Italy

FEATURES  
 source  
 1..1734  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..1734  
 /gene="TRIM15"  
 1..1398  
 /note="TRIM15"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="tripartite motif protein TRIM15 alpha"  
 /protein\_id="AA053505.1"  
 /db\_xref="GI:12407433"  
 /translation="MPATPSLKVHVELPACTLACGLPDAVITPGTTHPCRLPALP  
 OMGROSSKILLCLEEDQAPVPLGPIGPEYCEHGEKIVFCEBNAEPLC  
 VFCREBPQHOAHVYPLDEALOPRYRDLRSLTELSTREPEIDYKQEDOKLOVLT  
 OIESKHOVFAERLQELBQOCLLARIREBQIWKREDEYITKVSSEYTRILA  
 QVKELEKCOQPASELQDVVNQSRKEMTFSPEAISPQVYKIDFPHKLTITPE  
 MMRPFSNLNHLIEDSVITLDPQTSRSLVISEDKSVYTRQKSLDPSLPDPG  
 LPAVLGPFSSGSRHMVDLQLDGCGCTVYGVGEVRRKGEGLSAEDGVMAVVIS  
 HQOCWASTSPGTDLPLEIPRGVVALDYEAQVYTLNMQTQEIFFTFASFSKVPF  
 FFAWKRKGSCLTKG"

BASE COUNT 369 a 497 c 547 g 321 t  
 ORIGIN

Query Match 49.9%; Score 1301.4; DB 9; Length 1734;  
 Best Local Similarity 93.1%; Pred. No. 7.4e-285;  
 Matches 1630; Conservative 0; Mismatches 81; Indels 40; Gaps 24;

QY 527 ggtgtccatagagctgctgctgtacccctgtgtgagggcgctgagagatgctgtac 586  
 |||||  
 Db 24 GGTGTCATGAGCTGCTGCTGTACCTCTGTGCGGGCGCTGAGAGATGGCGTAC 83  
 QY 587 catccctgtggaacacacctctgcgggtctgctcccgagctctcccaatgagggc 646  
 |||||  
 Db 84 CATTCCTGTGAGACACACTTCTGCCGCTGTGCTCCCGCTCCAGATGGGGC 143  
 QY 647 ccaatccctgaggaagatctgctctgcggctctgcaagaagagtagagagagag 706  
 |||||  
 Db 144 CCAATCTCTG-GGCAAGATCTGCTCTGCCGCTGTGCCAAGAGAGAGAGAGAGAG 202  
 QY 707 ctcccatgccccctgtgccccctggcccgctggagataactactgtagagagagagc 766  
 |||||  
 Db 203 CTCCATGCGCCCTGTGCCCCGCGGCGGTGGGGA-AACTTACTGCGAGAGACACGCG 261  
 QY 767 gagaagatctactctctctgagagaagatgagagagagagagagagagagagag 826  
 |||||  
 Db 262 GAGAAGATCTACTTCTT-TGCGAGAAAGATGCCAGTCTCTGTGTCTGAGAGGA 320  
 QY 827 ggtcccaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 886  
 |||||  
 Db 321 GGGTCCACGACACAGGCGCACCGGTGGGTCTCTGACAGAGGCACTTACGCTACCG 380  
 QY 887 gtagctgtcagagtagctgagagagctctgagcagagagagagagagagagagag 946  
 |||||  
 Db 381 GGATGCTCAGAGAGTCACTGAGAGCTCTGAGCAGGAGAGAGATGATG-AGATG 439

OY	947	taaagtgtcaagaagaccgaagaactccaagtgtgggtgtgactcaagtctgaacaagaaga	1006
Db	440	TAAGTGTCAAGAAGACCACCAAGCTTCAAGTGTCTGACTGACTCAATCGA--AAGCAAA	497
OY	1007	gcccgtcaggtgtgcaacagctcctctgaagagctgtcaagcggagagctgcacgacgcat	1066
Db	498	GCATCAGGTGGAAACACACT---TTTATGAGAGCGTGC-AGCAGAGAGCTGGAGCGACGCAT	553
OY	1067	gtctccgtcttgaggccagactgagttggtacgctccggagtcacaaagtttggaaaggagag	1126
Db	554	GTCTCTGCTGTGGCCAGGCTAGGAGAGCT-----GGAGCAGCAGATTTTGAAGGAGAGGG	607
OY	1127	atgataatcacaaaggtctctcgaggaaagtcacccggctttggaocccagctcaagaag	1186
Db	608	ATGAATATATACAAAGGTCCTTGAGAAAGTCAACCGGCTTGAG-CCAGGTCAAGAG	666
OY	1187	ctcgagagagaagtgtaagcaagccagcaagtgaagctctcaagaatgtcagaagtcaagc	1246
Db	667	CT-GGAGAGAAAGTGTAGAGCCACCAAGTAGCTTTCACAAAGATGTACAGTCA--CC	724
OY	1247	agaagagtgtaagatbaaactttgtagtccttgaagcaattctccctbaactgttc	1306
Db	725	AGACAGGTGTGAATATAACACTTTTGTAGTCTTAGGCCATTTCT-CTTAGCTTGTCT	783
OY	1307	aagaagatccgtgatttccacaggaanaatactaacctccacagatatgatgaattgttc	1366
Db	784	AAGAAGATCCGTGATTTTCCACAGGAAATATCATCCCTCCAGAAATGATGAGCATGTTC	843
OY	1367	tcaagaanaacttggcgatcatcttgaaataagttgaagggtcatcatctcttgagacctca	1426
Db	844	TC-AGAAAACTTGGCGCATATCTGGAAATAGATTGAGGGGTATCACTGTGAGACCTCA	902
OY	1427	gaccgcagacgcggagacctgtgtctctcgaaagacaggaagtcagtgagttacacccgac	1486
Db	903	GACCGCCAGCCGGAG-CCTGTTCTCTCGGAADACAGAAATGAGATGAGATACACCCGC	961
OY	1487	agaagaagagcgtcgcaagaagccctctgctctgaagagctcccgcggttctcggct	1546
Db	962	AGAAAGAAAGCTCTCCAGACAGCCCTCGGTGCTTCACGCGCTCCGGGGGTCTGGGCT	1021
OY	1547	tcccgagctctctcccgagcgcaacgcgtctgacagttbaacttgagcttgagcgacgacg	1606
Db	1022	TCCCGGCTTCTCCTCCGGCGCCACCGGTGAGGTTACCTGCAGCGTGGCGGACGGCG	1081
OY	1607	gcggcgtcacaagtgtaggggtgtgccggggaggggtgagagacagagagatgtgagctc	1666
Db	1082	GGCGCTCACGGTGGGGGTGGCCGGGAGGGGTGAGGAGA-AGGAGAGATGTGGACTC	1140
OY	1667	agcgccagagacgcgtctcgagcgctgatalcatctctgcacaagaagatgcttgagccaca	1726
Db	1141	AGGCCCGAGAGCGCGTCTGGCCGTGATCATTC--GCACCAAGCATGTCTGGCCACGA	1198
OY	1727	ctctcccgagcacaacgaaactgtgccgctgagcagatccgcgcaagcgtgagatgtccct	1786
Db	1199	CCTCCCGGGCACCGACCTG-CCGCTGAGCAGATCCCGCGC-GGCGTGAAGATGTCCCT	1256
OY	1787	ggaatacgaagcgaggagtgagctctccaacagcccaagcccaaggggcccatctctc	1846
Db	1257	GGACTACGAGAGCGGGGAGGTGACCTTCACAAACGCCACGA--CCCAAGAGCCCATTTTC	1314
OY	1847	acctcaactgtctcttctcggccaagaagttctcccttctcttgccgcctctgacacaa	1906
Db	1315	ACCTTCACTGCTCTT--CTCCGGCAAAAGTCTCCCTTCTTTTGCCTGTG---AAAA	1369
OY	1907	agggtcctgacctgaagctgtgacacgagggaaatbgggcgcgcaaggcgcgcaagcgga	1966
Db	1370	AAGGTCTCTCTTACCTGA- AAGGCTGAAGTGGGGCGCGGAAGGGCGGCGAAGCGGA	1428
OY	1967	gagcggcgactctcggagatccaactctcgccctctgcgcagtgtagcgccgggggtccct	2026
Db	1429	GACGGCGGCTCTCCGGGATCACTCTCGCCCTCGGCGCAATGTGCGCGCCCGGGGCTCCT	1488

Accession	Sequence	Length
OY 2027	gtgcgcgcgtgaagcgagagaaacacgagacttgactctgaacacgagctgtttttac	2086
Db 1489	gtgcgcgcgtgaagcgcagaaA--CAGGGGACTTGAGTGTGCAACAGCGGTTGTTTAC	1546
OY 2087	ttatattatcttaggccctcagctctccctgaagtcctgaagcctccctgtgaagctctgcgc	2146
Db 1547	TTTATTATCTTAGGCCCTCAGCTCCCTCAGCTTCGAGCCTCCCTGTGAGCGCTTGGCC	1606
OY 2147	ttctctgcacctaaagatgtagaacacacagacgacttgagctgctgctgctctaaaggaaacagcc	2206
Db 1607	TTCTCTGCACCTCAGAGTGTGAGAACACAGACGCGCTTCGGCTGTGCTTAAGGGCAACAGCC	1666
OY 2207	aacctagaagaccgcgcgcctcttcgcyggaanaactaaagagagacatctaaatgtaa	2266
Db 1667	AACCTAGAGACCCAGCGGCTTTTCGGGAAAAA---AAAGAAAAACACTTAATAATAAA	1723
OY 2267	tgtttaactg 2277	
Db 1724	TGTTTAAC TG 1734	

LOCUS	AC004189/c	RESULT 3
DEFINITION	Homo sapiens clone UMC:Y18c247 from 6p21, complete sequence.	PRI 17-JUL-1998
ACCESSION	AC004189	
VERSION	AC004189.1	
KEYWORDS	HTG.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 47777)	
JOURNAL	Jener.M.M., Guillaudeau,T., Vu.Q., Kutayavl.N.T., Harter,H. and Geraghty,D.E.	
REMARK	Large scale sequence analysis of the human MHC class I region Unpublished (1998)	
REFERENCE	Fred Hutchinson Cancer Research Center	
AUTHORS	The Clinical Research Division	
TITLE	1100 Fairview Ave. N., P.O. Box 19024	
JOURNAL	Seattle, WA 98109-1024	
REFERENCE	2 (bases 1 to 47777)	
AUTHORS	Geraghty,D.E. and Olson,M.V.	
TITLE	Direct Submission	
JOURNAL	Submitted (23-FEB-1998) Human Genome Center, University of	
REFERENCE	Washington, Box 352145, Seattle, WA 98195, USA	
AUTHORS	3 (bases 1 to 47777)	
TITLE	Geraghty,D.E. and Olson,M.V.	
JOURNAL	Direct Submission	
REMARK	Submitted (17-JUL-1998) Human Genome Center, University of	
COMMENT	Washington, Box 352145, Seattle, WA 98195, USA Contact: Daniel E. Geraghty (geraghty@fhcc.org) On Jul 17, 1998 this sequence version replaced gl:2905871. Overlapping Sequences: 5' : UMG:Y17c054 (genbank Accession: AC004186) 3' : UMG:Y2c224 (genbank Accession: AC004202)	

Sequence Quality Assessment: This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Double stranded (DS) coverage:	88.8
DS or two chemistry coverage:	99.0
Single stranded regions:	2

This sequence has been validated by Multiple Complete Digest sequence validation.



OY	1823	ccagagcccaaggcccaatccctcaacctcaatctgacctctctctccgagcaagttctcc	1882
Db	15728	CCAGA--CCCAAGAGCCCATTTCACTTCACTGCTCTTT--CTCCGGCAAAAGTCTTCC	15673
OY	1883	ctgtctctctgcccgcctctgacacaagaaggtctccctgacctagctctgacacggggaaatg	1942
Db	15672	CTTTCTTTGGCGTCTGG---AAAAAAGGTTCCTGCTTACGCTGA--AAGGCTGAAGTGGG	15617
OY	1943	gcgcgcgaagggcggcgcgaagcgcgaagcggcgctctctccggatccagctccgcctctgc	2002
Db	15616	GGCGCCGGAAGGCGCGCGAAGCGGAGACGCGCGCTCTCCGGGATCCACCTCCGCCCTTGGC	15557
OY	2003	cagttgtgagcccccgggggtctccctctgtcccgctggaagcgagagaacaagggtactga	2062
Db	15556	CAGTGTGGGGCCGGGGGGCTCCCTGTGGCCCGCGTAGGCGAGAGAAACA--GGGGACTTGA	15499
OY	2063	gtctcgaacagcggctgtcttcttaacttlaattatcttagccctcagctccctgacgtcct	2122
Db	15498	GTCTCGAACAGCGGGTGTCTTTACTTATTTATCTTAGGCCCTCAGCTCCTTGACGTCT	15439
OY	2123	gagctccctgtgacgtctctgacctctctctctgacacctagatgcagacaacagagct	2182
Db	15438	GAGCTCTCCCTGTGACGCTCTGGCTTCTCTCTCACCCTGACAGATGACAGAACACACACGCGCT	15379
OY	2183	tcggtctgtgaccttaaggagacagccaaccttagaacccgcgcgcctctccggggaaaaacta	2242
Db	15378	TGCGGTGTGGCCTAGGGGCAACGCCAACCTTAGAGACCAAGCGGGCTTTGGGGAAAAA--A	15322
OY	2243	aagaagagacatcctaaatgttaatgtttaactgcttcaagataatctatcttggaaaa	2302
Db	15321	AAGAAAGACATCTAAAAATAAATGTTTAAACGTGTTCAAAATATATTTCTTGGAAAA	15262
OY	2303	atcaaggtcttgcctgagcttgcactaatgtgtacagttacactctcgtaactctgtacacacac	2362
Db	15261	ATCAGGGTTTGGCTGGAGACTTCACATAATTTGTACAGTTAACTTGTACTTTGACACACAC	15202
OY	2363	ctgaaatagctctcccaacctctttagagcttagagccttttaacagccctgggtgagacc	2422
Db	15201	CTGAAGATGCTCCACACCTTGTAGGGCTTAGGGCTTTTATACGCCCTGGGTGGACCCC	15142
OY	2423	agggcacctctcttccctccctccctctgtgtaacttctctgtgactctttagaagaatgctcta	2482
Db	15141	AGGGCCCTTCTCTTCCCTTCCCTTCTGCTATTTCTCTGAGCTTGTAGAGAAATGCTCTTA	15082
OY	2483	agaaagtgtgactcaacagacctctgtagtccatgtgttccaaatagtcgctatgaggtactg	2542
Db	15081	AGAAAGTGTGACTCACAGACCTGTGATTTGCATGTGTCCAAATPAGCCTGATGGAGACTGG	15022
OY	2543	agaaaggtctaataatccaatgagatctgtgcctgtgttgcaattagaagccgagatgtctc	2602
Db	15021	AGAAAGGCTTTAATTCATGGGATTC--TCCGTGTGTGGCAATTTAGGCGCGAGATGGCTTC	14963
OY	2603	gagagagat 2610	
Db	14962	GAGGAGT 14955	
RESULT 4	AP000517/c	AP000517 100000 bp DNA	PRI 30-MAR-2000
LOCUS	DEFINITION	Homo sapiens genomic DNA, chromosome 6p21.3, HLA Class I region, section 16/20.	
ACCESSION	AP000517	AP000517	
VERSION	AP000517.1	GI:5926704	
KEYWORDS	SOURCE	Homo sapiens DNA.	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.	
AUTHORS		Shima,S., Tamiya,G., Oka,A. and Inoko,H.	
TITLE		Homo sapiens 2,729,817bp genomic DNA of 6p21.3 HLA class I region	
JOURNAL		Published Only in Databases(1999) In press	

REFERENCE 2 (bases 1 to 100000)  
AUTHORS Hirakawa, M., Yamaguchi, H., Imai, K. and Shinada, J.  
TITLE Direct Submission  
JOURNAL Submitted (21-Sep-1999) to the DDBJ/EMBL/GenBank databases. Mita  
Hirakawa, Japan Science and Technology Corporation (JST), Advanced  
Databases Department; 5-3, Yonbancho, Chiyoda-ku, Tokyo 102-0081,  
Japan (E-mail: mika@tokyo.jst.go.jp,  
URL: http://www-alis.tokyo.jst.go.jp/, Tel: 81-3-5214-8491,  
Fax: 81-3-5214-8470)  
COMMENT This sequence is conducted by Tokai University as a JST sequencing  
Team.  
Principal Investigator: Hidetoshi Inoko Ph.D  
Phone: +81-463-93-1121, Fax: +81-463-94-8884.  
The sequence is submitted by Human Genome Sequencing in ALIS  
Project of JST  
Japan Science and Technology Corporation (JST)  
5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan  
For further information about this sequence, please visit our  
sequence archive Web site (<http://www-alis.tokyo.jst.go.jp/RGS/top.html>) or send email to [webmaster@www-alis.tokyo.jst.go.jp](mailto:webmaster@www-alis.tokyo.jst.go.jp).  
FEATURES  
SOURCE Location/Qualifiers  
1. .100000

Query Match	Best Local Similarity	Score	DB 9	Length
Matches 1140: Conservative	94.4%	Pred. No. 3.2e-210;	Mismatches 48;	Indels 20; Gaps 12;
Db 70311	AGGGGTCACTCACTCGGACCCCTCAACACCGCCGCGAG-CCTGGTTCTCTCGAAGACA	1403	aggagtcacaccttgagacctgagacctgagccgagccgagagagctgtctctcgaagaca	1462
Db 70252	GGAACTCAGTAGGACCCGGCGAAGAGAGAGCTGCGACACGCCCTTCCGCTTCC	1463	ggaagtcagtagagtagaacaccgagcgaagaagaagagctcgaagaagagcccccgcgcttcg	1522
Db 70252	GGAACTCAGTAGGACCCGGCGAAGAGAGAGCTGCGACACGCCCTTCCGCTTCC	1463	ggaagtcagtagagtagaacaccgagcgaagaagaagagctcgaagaagagcccccgcgcttcg	1522
STS	complement(38199..38348)		/standard_name="D6S2198"	
STS	/note="SHGC-16144:The location is between each flanking site of PCR primers."		/db_xref="GDB:741200"	
STS	complement(45322..45412)		/note="9065:The location is between each flanking site of PCR primers."	
STS	/db_xref="GDB:1297693"		complement(52370..52547)	
STS	/standard_name="D6S2267"		/note="sy924b3-L:The location is between each flanking site of PCR primers."	
STS	/db_xref="GDB:5583975"		57335..57660	
STS	/standard_name="D6S1944"		/note="WI-7404:The location is between each flanking site of PCR primers."	
STS	/db_xref="GDB:679168"		57336..57491	
STS	/note="RH17987:The location is between each flanking site of PCR primers."		/db_xref="GDB:4572735"	
STS	complement(81490..81630)		/note="SHGC-17229:The location is between each flanking site of PCR primers."	
STS	/db_xref="GDB:1231045"		complement(88508..88638)	
STS	/note="SGC34813:The location is between each flanking site of PCR primers."		/db_xref="GDB:4584664"	
BASE COUNT	28701 a 22059 c 22429 g 26811 t			
ORIGIN				

QY 1523 acgagctccgaggtctctggtctccgggctctcccgagcgccacgctggcagg 1582  
 |||||  
 Db 70192 ACGGCTCCCGGCGGTCTGGGCTTCCGGGCTTCTCCGCGGCGCCACGCGTGGCAGG 70133  
 |||||  
 QY 1583 ttgacctcagctggcgagcgagcgagctgacaggtgggggtggcgaggaggtgca 1642  
 |||||  
 Db 70132 TTGACCTCAGCTGGGCGACGCGCGGCTGACGCTGGGGTGGCGCGGAGGGGCTGA 70073  
 |||||  
 QY 1643 ggaagacagagagatgggactacgagcgagagagcggtctggcgctgatactct 1702  
 |||||  
 Db 70072 GGAGGA-AGGAGAGATGGGACTACGCCCGAGAGGCGGCTGGGCGGTGATCATCTCG 70014  
 |||||  
 QY 1703 gacccaagcagtgctgggccaagcactcccgagcaccgactgctgcgtgagatc 1762  
 |||||  
 Db 70013 CACG--ACGAGTGTGGGCGACGACCTCCCGGGCACCAGACTG-CCGCTAGCGAGATC 69957  
 |||||  
 QY 1763 ccgagcaggtgagatgcgcctggactacgagcgagggcgaggtgacccctcacacgc 1822  
 |||||  
 Db 69956 CCGGCG-GGCGTGAAGTCCGCTTGACTACGAGCGGGGCGAGGTGACCTCCACACGCG 69898  
 |||||  
 QY 1823 ccaagagccaggggccaactcctcaactgctcttctccggccaagctctcc 1882  
 |||||  
 Db 69887 CCAGG--CCGAGAGCCATCTCACTCACTGCTCTT--CTCCGGCAAGGCTTCC 69842  
 |||||  
 QY 1883 ctgtccttgagcgcctggacacaaagggtccctgacctagagctgacagcgaggaaatgg 1942  
 |||||  
 Db 69841 CTTCTTTGGCGGTGCG--AAAAAGGTTCTCGCTTACGCTGA-AAGGCTGAAGTGG 69786  
 |||||  
 QY 1943 ggcgcgagagggcgagcagcgagagcgagctccgggataccagctccgcccctg 2002  
 |||||  
 Db 69785 GCGCGCGAAGGCGCGCGAAGCGAGACGCGGCTCCGCGATCCAGCTCCGCCCTGGC 69726  
 |||||  
 QY 2003 caggtgtggcgccggggggtccctgtgcccggcgatgagcgagagaaacagggagctga 2062  
 |||||  
 Db 69725 CAGGTGGCGCCGGGGGCTCCCTGTGCGCGGTAGCGAGAGACA--GGGCGCTTGA 69668  
 |||||  
 QY 2063 gctcgacaacaggtgtgttcttacttacttacttacttacttacttacttacttactt 2122  
 |||||  
 Db 69667 GTCGCGAAGCAGCGGTGTTTACTTATTTATCTTAGGCCCTCAGCTCCGAGAGTCTCT 69608  
 |||||  
 QY 2123 gagctccctgtgagctctgctctctctgacactcagatgacagaaacagagct 2182  
 |||||  
 Db 69607 GAGCTCTCCCTGAGCGCTCTGCTCTCTGCTGACCTCAGAGACGAGACACAGCGGCT 69548  
 |||||  
 QY 2183 tcgagctgtgcttaggggcaagcgcaagccttagaagcggcgcccttggggaaacta 2242  
 |||||  
 Db 69547 TCGGCTGTGCTTAGGGCAACGCAACTAGAGCGAGCGGGCTTGGGGGAAAA--A 69491  
 |||||  
 QY 2243 aagaagagacatctaaatgtlaatgtttaaactgttccaagataatactctgggaaaa 2302  
 |||||  
 Db 69490 AAGAAAGACATCTAAATTAATTAATGTTAACTGTTCAAAATTAATTAATCTTGGGAAA 69431  
 |||||  
 QY 2303 atcaggttctgctgagctctgacttaattgtacagttacttacttacttacttactt 2362  
 |||||  
 Db 69430 ATCAGGGGTTTGTGAGACTTGTGACTAATTTGACAGTTACTTCTACTTGTGACACAC 69371  
 |||||  
 QY 2363 ctgaagatgctcaccctttagagcttagggcctttagcagccttggtagagccc 2422  
 |||||  
 Db 69370 CTGAAGATGCTCAGCTTGTGAGGCTTAGGGCTTATTTATCAGCCCTGGGTGAGCCG 69311  
 |||||  
 QY 2423 agggcccccttctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2482  
 |||||  
 Db 69310 AGGGCCCCCTTCTTCCCTTCCCTTCTGCTGATTTCTGTGAGCTGTGAGAGATCTCTA 69251  
 |||||  
 QY 2483 agaaagtgtgactacagacactctgagatccaatgtccaatlagcgctgagagctgg 2542  
 |||||  
 Db 69250 AGAAGGTGATCTACACACCTGTGATTCATGTCTCAATTAACGCTGATGAGGACTGG 69191  
 |||||  
 QY 2543 aagaagcttaatccaatggatctgctgtgtgtgcaatlttggcgagagatgagctc 2602  
 |||||  
 Db 69190 AGAAGGCTTAATTCATGGGATC-TGCTGTGTGGCAATTTAGGGCGAGATGGCTC 69132  
 |||||  
 QY 2603 gagggagt 2610

Db 69131 GAGGAGT 69124  
 |||||  
 RESULT 5  
 AB023054/c  
 LOCUS  
 DEFINITION  
 Homo sapiens genomic DNA, chromosome 6p21.3, HLA class I region,  
 clone:679622, complete sequence.  
 ACCESSION  
 AB023054  
 VERSION  
 AB023054.1  
 KEYWORDS  
 HTG.  
 SOURCE  
 Homo sapiens  
 ORGANISM  
 Homo sapiens DNA, clone:679622.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (sites)  
 Shihina,T., Tamiya,G., Oka,A., Takishima,N., Yamagata,T.,  
 Kikkawa,E., Iwata,K., Tomizawa,M., Okuaki,N., Kuwano,Y.,  
 Watanabe,K., Fukuzumi,Y., Itakura,S., Sugawara,C., Ono,A.,  
 Yamazaki,M., Tashiro,H., Ando,A., Ikemura,T., Soeda,E., Kimura,M.,  
 Bahram,S. and Inoko,H.  
 Molecular dynamics of MHC genesis unraveled by sequence analysis of  
 the 1,796,938-bp HLA class I region  
 Proc. Natl. Acad. Sci. U.S.A. 96 (23), 13282-13287 (1999)  
 20027539  
 JOURNAL  
 MEDLINE  
 REFERENCE  
 2 (bases 1 to 162383)  
 Shihina,T. and Takishima,N.  
 Direct Submission  
 Submitted (29-JAN-1999) to the DDBJ/EMBL/GenBank databases. Takashi  
 JOURNAL  
 Shihina, Tokai University School of Medicine, Department of  
 Molecular Life Science 2; Bohseidai, Isehara, Kanagawa 259-1193,  
 Japan (E-mail:tsuhina@is.icc.u-tokai.ac.jp, Tel:81-463-93-1121,  
 Fax:81-463-94-8844)  
 Location/Qualifiers  
 FEATURES  
 source  
 1..162383  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="679622"  
 /map="6p21.3"  
 /map="6p21.3"  
 BASE COUNT 45921 a 36561 c 36078 g 43823 t  
 ORIGIN  
 Query Match 37.2%, Score 971.2; DB 9; Length 162383;  
 Best Local Similarity 94.4%; Pred. No. 2.9e-210;  
 Matches 1140; Conservative 0; Mismatches 48; Indels 20; Gaps 12;  
 QY 1403 aggggtcaactctggagccctcagacggcagcgagagctgttctctcgagagca 1462  
 |||||  
 Db 63731 AAGGCTATCACTGTGGACCTCAGACCGCCAGCGGAG-CTGTGTTCTTGGAGACA 63673  
 |||||  
 QY 1463 ggaagtcagtgagtlacacccgagcagaagaagcctgcagacagacccctgcctcg 1522  
 |||||  
 Db 63672 GGAAGTCAGTGAAGTACACCGCGGAGAGAGAGCTCCAGACAGCCCTCGGCTTCG 63613  
 |||||  
 QY 1523 acggctccggcggtctctgggtctccgggctctcctcgcggcgacagcgtgcaag 1582  
 |||||  
 Db 63612 ACGGCTCCCGCGGCTTGGGCTTCCGGGCTTCTCTCCGCGGCCACCGCTGGCAGG 63553  
 |||||  
 QY 1583 ttgacctcagctgggagcagcgcgagctcagagtgagggtggcgaggaggtgga 1642  
 |||||  
 Db 63552 TTGACCTCAGCTGGGCGACAGCGCGGCTGACGCTGGGGTGGCGGGGAGTGA 63493  
 |||||  
 QY 1643 ggaagacagagagatggagctcagcgccgagagcagcgctgtggcgctgatactct 1702  
 |||||  
 Db 63492 GGAGGA-AGGAGAGATGGGACTCAGCCCGAGAGCGGCGTGTGGCGGTATCATCTCG 63434  
 |||||  
 QY 1703 gacccaagcagtgctgggacagacactcccgggacacgagactgtcgtctgagagatc 1762  
 |||||  
 Db 63433 CACG--ACGAGTGTGGGCGACGACACCTCCGCGGCGACGAGACTG-CCGCTGAGCGAGATC 63377

[illegible][illegible]

QY 2003 cagctgctgagcccgagccctcctgctgcccgcgtgagcgagagaacacgagacttga 2062  
|||||  
Db 63145 CAGTGTGCGGCGCCGGGCGCTCCCTGTCGCCGCGTGAGGGGAGAGAAACA--GGGGAGCTTGA 63088  
QY 2063 gctctgagaaagaggtgtgttttacttatttctctcttaggcctcagctccctgaagtcct 2122  
|||||  
Db 63087 GCTCTGAAACAGGGGCTGTTTACTTATTTATCTTAAAGCCCTCAGACTCCCTGACGTCCT 63028  
QY 2123 gagctccctgtagagctctgctctctctctgacacctcagaatgacacacagagagct 2182  
|||||  
Db 63027 GAGCCTCCCTGAGAGCTCTGCGCTTCTCTGACACTCAGAGTCGAGAACACAGAGGCT 62968  
QY 2183 tcgagctgtagcttagagcaacagcaacacttaggaacccgcgcctctcgggagaaaaacta 2242  
|||||  
Db 62967 TCGGCTGTGCTCCTAGGGCAACACCACTAGAGCCAGCGGCTTTCGGGAGAAAA--A 62911  
QY 2243 aagagagagacatctaaatgtatgtttaaactgttccaagataattactcttgagaaaa 2302  
|||||  
Db 62910 AAGAAAAAGACATCTAAATAATATAATGTTTAACTGTTTCAAAATATATATCTTGGGAAAA 62851  
QY 2303 atcagagtttctgtagctgtagctgtagctgtagctgtagctgtagctgtagctgtagctg 2362  
|||||  
Db 62850 ATCAGAGGTTTCTGACTGTGACTGATTTGACAGTTACTTCTGACTTTGACACACAC 62791  
QY 2363 ctgaagatgctcccaacctgtgtaggagcttaggagcttctatcagccctggtgagcccc 2422  
|||||  
Db 62790 CTGAAGATGCTCCACACTTTGTAGGGCTTAGGGCTTTTATCAGGCCCTGGGTGACCCC 62731  
QY 2423 agggccctctctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 2482  
|||||  
Db 62730 AGGGCCCT 62671  
QY 2483 aagaagtgtagctgtagctgtagctgtagctgtagctgtagctgtagctgtagctgtagctg 2542  
|||||  
Db 62670 AAGAAAGTGACTACAGACCTCTGATTCAGTGTCCAAATTAAGCGCTGATGGGACTG 62611  
QY 2543 aagaagcttaacatcaatgagatctgctgtgtgtgcaatttgagggcgagatgagctc 2602  
|||||  
Db 62610 AAAAAAGCTTAATCCAATGGGATC-TGCGTGTGTGGCAATTTAGGCGGAGATGCTC 62552  
QY 2603 gagggagct 2610  
|||||  
Db 62551 GAGGAGT 62544

RESULT 7  
HSRINGG/c 17284 bp DNA PRI 05-AUG-1999  
LOCUS HSRINGG Homo sapiens RFB30 gene for RING finger protein.  
DEFINITION Y07829  
ACCESSION Y07829.2 GI:5708209  
KEYWORDS major histocompatibility complex; RFB30 gene; RING finger protein.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 17284)  
Tazai-Almihl, R. and Pontarotti, P.  
Henry, J., Ribouchon, M., Depetris, D., Mattei, M., Offer, C.,  
Cloning, structural analysis, and mapping of the B30 and B7  
multigenic families to the major histocompatibility complex (MHC)  
and other chromosomal regions  
Immunogenetics 46 (5), 383-395 (1997)  
JOURNAL MEDLINE  
REFERENCE 97419168  
AUTHORS Pontarotti, P.  
TITLE Direct Submission  
JOURNAL Submitted (06-SEP-1996) P. Pontarotti, Unite 119 INSERM, 27 bd. Lei  
Route, 13009 Marseille, FRANCE  
REMARK Revised by [4]  
REFERENCE 3 (bases 1 to 17284)  
AUTHORS Pontarotti, P.  
TITLE Direct Submission

JOURNAL Submitted (04-AUG-1999) P. Pontarotti, Unite 119 INSERM, 27 bd. Lei  
Route, 13009 Marseille, FRANCE  
COMMENT On Aug 6, 1999 this sequence version replaced gi:1770497.  
FEATURES  
source location/Qualifiers  
1. 17284  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.3"  
/tissue\_type="testis"  
914. 1258  
/gene="RFB30"  
/note="Exon A1"  
join(914. 1258,8899. 9327,9338. 10512,10513. 10608.  
10609. 11125,11126. 11356,11357. 12358,12359. 12381,  
12382. 12696,12697. 12812,12813. 13987,13988. 14020,  
14021. 15255,15256. 15773)  
/gene="RFB30"  
complement(2228. 2583)  
/note="L1M4"  
/rpt\_family="LINE/L1"  
complement(3067. 3217)  
/note="L2"  
/rpt\_family="LINE/L2"  
complement(3218. 3485)  
/note="AluSg1"  
/rpt\_family="SINE/Alu"  
complement(3486. 3623)  
/note="L2"  
/rpt\_family="LINE/L2"  
3754. 4608  
/note="L1PA7"  
/rpt\_family="LINE/L1"  
complement(4675. 5170)  
/note="L2"  
/rpt\_family="LINE/L2"  
complement(5655. 6042)  
/gene="RFB30"  
complement(5655. 6042)  
/gene="RFB30"  
/note="Exon A2"  
complement(6910. 7048)  
/note="MIR"  
/rpt\_family="SINE/MIR"  
7308. 7409  
/note="MER33"  
/rpt\_family="DNA/MER1\_type"  
complement(7433. 7741)  
/note="AluY"  
/rpt\_family="SINE/Alu"  
7743. 7944  
/note="MER33"  
/rpt\_family="DNA/MER1\_type"  
8041. 8178  
/note="MIR"  
/rpt\_family="SINE/MIR"  
8899. 9327  
/gene="RFB30"  
/number=1  
join(8899. 9327,10513. 10608,11126. 11356,12359. 12381,  
12697. 12812,13988. 14020,15256. 15773)  
/gene="RFB30"  
/codon\_start=1  
/product="RING finger protein"  
/protein\_id="CAB52384.1"  
/db\_xref="GI:5708210"  
/translation="MASAASVTSIADENVNCPICQGLREPVITDCGHNFCRACTLRVC  
EIPGDLDESPCLCKEPFSPFPMWOLANVETRLQVSTLIGEDVVOEH  
GKIYFCEDEDMQCYCRAGAHATHTMFLDAAAPYRQIHKCKLRKEEI  
DEIOSRENNQVLTQVSTRQOVTSFAHRLKFLDEQSLILAOLESQGDILKOR  
DEFDLVAGETICRESALIIELEEKNERARELTDIISTLIRCTRKCRKPAVAPSL  
GORIDPQOALPLQREMKFLEKICFELIDYEPAHISIDPQTSHEKILISEDHQAO  
SSKQNSPDNPFQDRATCAVLAHTGISTGRHTWVYSIDLHAGASCTGVAVSEVDVRK

```
ELRLRPEGVWAVRLAMGFVSAIGSFPTRLTLKEQPROVNSLDYEVGWTFNNATVR
EPIYTFSTFTRKVIPIPFGLMGKSSFSISS"
intron
9338..10512
/feature="RFB30"
/number=1
repeat_region
9521..9570
/feature="L2"
/feature="LINE/L2"
exon
10513..10608
/feature="RFB30"
/number=2
intron
10609..11125
/feature="RFB30"
/number=2
repeat_region
10690..10886
/feature="MIR"
/feature="SINE/MIR"
exon
11126..11356
/feature="RFB30"
/number=3
intron
11357..12358
/feature="RFB30"
/number=3
repeat_region
12039..12276
/feature="AluA0"
/feature="SINE/Alu"
exon
12359..12381
/feature="RFB30"
/number=4
intron
12382..12696
/feature="RFB30"
/number=4
exon
12697..12812
/feature="RFB30"
/number=5
intron
12813..13987
/feature="RFB30"
/number=5
repeat_region
12856..13010
/feature="MIR"
/feature="SINE/MIR"
complement(13604..13715)
repeat_region
13709..13857
/feature="LINE/L2"
repeat_region
13988..14020
/feature="DNA/MER1_type"
exon
14021..15255
/feature="RFB30"
/number=6
intron
15256..15773
/feature="RFB30"
repeat_region
15084..15084
/feature="MER5A"
repeat_region
15088..15129
/feature="MER5A"
exon
15256..15773
/feature="DNA/MER1_type"
repeat_region
16057..16163
/feature="MIR"
/feature="SINE/MIR"
BASE COUNT 4637 a 3801 c 4162 g 4586 t 98 others
ORIGIN
```

Query Match 27.4%; Score 714.4; DB 9; Length 17284;  
Best Local Similarity 91.9%; Pred. No. 7.6e-152;  
Matches 817; Conservative 0; Mismatches 38; Indels 34; Gaps 5;

```
QY 1 gatgtggcagcgcctcagagcagaagttatggctccaccctgctcaatctgacgaag 60
|||||
Db 6513 GCTGTGGGACGACCTCAGAGCAGAAATTATGCTCCACCTGCTCAATCTGACAGAA 6454
QY 61 gctctgtctcccaagctctcccaagcactgtgtgttaacagattccaggaacccatccc 120
|||||
Db 6453 GCTTCTGCTCCCAAGTCTCCCAAGCCACTGTGTGATGATTCAGAAACCATCC 6394
QY 121 cctgtgacctcaggtgtgtctgtcttctccacctagaagcagaagaagcagaagtaa 180
|||||
Db 6393 CCTGTGACCTGATGCTGTGCTGCTTCTCCACCTAGAGACAGAAAGACGACGATTA 6334
QY 181 agaactgtgcttaacttgcgcgcacactgggaattctgggttaattcagaacgcctgaatt 240
|||||
Db 6333 AGAAGTGGCTTACTTGTGCGCCGACCTGGGAATTCGTGGTAATTCAGACCCCTNNNN 6274
QY 241 tggacccactccgctgtagttagtggcagaggtctcagaagaaacagaagcaggaacca 300
|||||
Db 6273 TGGACCACTCCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 6214
QY 301 ggtgtctccctgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 360
|||||
Db 6213 GGTGCTTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6176
QY 361 ctctctctctctctctctctcagccttgagccgcttccctccctcctgctcagtgctg 420
|||||
Db 6175 CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 6121
QY 421 agtgtgactcgaattcagaagaaagaaactcgcgtggcgtcagaagaaacaggaatggacg 480
|||||
Db 6120 AGTGTGACTCGATTTCAGAGGAAGGAAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6061
QY 481 gctgggaaagcagcgtgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 540
|||||
Db 6060 GCTGGGAAAGCAGCAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6006
QY 541 tgcctgcttgaacctctgtgctgctgctgctgctgctgctgctgctgctgctgctgctg 600
|||||
Db 6005 TGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5946
QY 601 aacactctcccgctctgctctcccgctcctcctcagatggagggcccaatcctcgtgac 660
|||||
Db 5945 ACACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5887
QY 661 aagatcctgctctgcccgcctctgccaagagagtagcagagcagagatccatgagccct 720
|||||
Db 5886 AAGATCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5827
QY 721 gtgcacctggccgcctgtagagataactactcgcgaagacagcgagagaagatctact 780
|||||
Db 5826 GTGCCCTGCGGCGCGCGGAGGAACCTTACTGCGAGAGACACGAGCAATCTACTT 5767
QY 781 ctctctgcagaagatgctcgaagctcctctgtgtgtctcagaagaggtcccaagcagc 840
|||||
Db 5766 CTTC-TGCGAAGACGATGCGAGTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5708
QY 841 aggcgcacacgctgggtctcctgagcagagccatcagccctcagcagga 889
|||||
Db 5707 AGGGGACACACGCTGGGTTCTCTGACGACGACATTCAGCCACCGGTA 5659
```

RESULT 8  
SSC251829  
LOCUS SSC251829 152211 bp DNA MAM 19-JUN-2001  
DEFINITION Sus scrofa MHC class I SLA genomic region, haplotype H01, clone BAC  
ACCESSION AJ251829 AF074423 Z97378 Z97379 Z97380 Z97403  
VERSION AJ251829.1 GI:6625535  
KEYWORDS acid finger protein; aip gene; major histocompatibility complex;  
MHC class I antigen; RFB30 gene; rfp gene; ring finger protein B30;  
sla- gene; swine leucocyte antigen; zinc finger protein; znfl73  
pig. znfl7 gene.



ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
REFERENCE 1 (bases 1 to 152211)  
AUTHORS Renard,C., Vaiman,M., Chianiliculchai,N., Catolico,L., Robert,C.  
and Chardon,P.  
TITLE Sequence of the swine major histocompatibility region containing  
the classical class I genes  
JOURNAL Immunogenetics In press  
REFERENCE 2 (bases 1 to 152211)  
AUTHORS Renard,C.  
TITLE Direct Submission  
JOURNAL Submitted (15-DEC-1999) Renard C., Animal Genetics, Inra Cea, Lre9,  
Domaine De Vilvert, Jouy-en-Josas, 78350, FRANCE  
COMMENT On Jun 25, 2001 this sequence version replaced gi:3328033  
gi:2252570 gi:2252572 gi:2252611.  
Submitted jointly by Dr C. Renard, (address as above) and Dr  
L.Catolico,  
(Genoscope, 2 rue Gaston Cremieux, PB191, Evry 91006, France).  
Sequence  
quality assessment: This entry has been annotated with sequence  
assembled  
by the Phrap program. 98% of bases, with quality above 40 as  
determined  
by the base-calling Phred program, has been confirmed either by two  
chemistry coverage on one strand, or at least 3 sequences  
determined on  
both strands.  
The 3' end of AJ251829 sequence (BAC 207G8) is followed by the 5' end  
of the AJ13112 sequence (BAC 490B10) after the position 33.  
FEATURES  
source  
1. 152211  
/organism="Sus scrofa"  
/strain="large white"  
/db\_xref="taxon:9823"  
/chromosome="7"  
/map="7p1.1"  
/clone="BAC 207G8"  
/haplotype="H01"  
/sex="male"  
/cell\_type="fibroblaste"  
/clone\_id="BAC"  
2148. 2430  
/rpt\_family="sine"  
complement(2611. 3170)  
/rpt\_family="line"  
complement(3267. 3560)  
/rpt\_family="sine"  
3293. 3316  
/note="(TTG)8"  
/rpt\_family="microsatellite"  
4791. 5038  
/rpt\_family="sine"  
7030. 7060  
/note="(AAT)9"  
/rpt\_family="microsatellite"  
complement(9849. 10126)  
/rpt\_family="sine"  
complement(12368. 12624)  
/rpt\_family="sine"  
complement(12776. 13050)  
/rpt\_family="sine"  
16349. 16649  
/rpt\_family="sine"  
16596. 16621  
/note="(CAA)8"  
/rpt\_family="microsatellite"  
complement(17721. 17975)  
/rpt\_family="sine"  
complement(20021. 20184)  
/rpt\_family="line"  
21343. 21357  
/note="(TA)7"

repeat\_unit  
complement(21368. 21620)  
/rpt\_family="sine"  
23560. 235904  
/gene="rfp"  
/number=1  
23560. 23904  
/gene="rfp"  
23560. 23904  
/gene="rfp"  
23560. 23904  
/gene="rfp"  
/codon\_start=1  
/pseudo  
/product="ring finger protein"  
26361. 26631  
/rpt\_family="sine"  
31071. 31090  
/note="(CAAA)4"  
/rpt\_family="microsatellite"  
complement(34434. 34439)  
/gene="rfb30"  
complement(34434. 48510)  
/gene="rfb30"  
complement(36573. 36862)  
/gene="rfb30"  
/rpt\_family="sine"  
complement(join(38204. 38721,40666. 40698,42436. 42551,  
43384. 43406,44457. 44687,45259. 45354,48079. 48510))  
/gene="rfb30"  
/codon\_start=1  
/product="putative ring finger protein B30"  
/protein\_id="CAB63932.1"  
/db\_xref="GI:6625536"  
/db\_xref="SPTREMBL:O19085"  
/translation="MASASVSSSLADEVNCPCICGTLREPYTIDCGHNFCCVCLTRYL  
EIPCIDGELPTCPICKEPRRGSFRPNWQLASVVENIERKLGSOLGSEEBEEVCLIE  
HREKYYVEEDDEMQLCVCREAMHHHTVRFLEDAAGPYREOIOKCLCTLRKEE  
IORIOLRENNORIOVLNOVATRKOKVISEFAHLISQPLIEOOSVLAOLERIDGILKH  
RDERRDVLVAGETICFNLTIELLEKNRPRADLTIDRSTLIRETRRCRPRAVSPE  
LGQRIRDPQOQALPLRRDKMTLEKLCFELDYEPANHSLDPQISHPKLLISEDNOQR  
FSYKWNQSPDNPORFRDRAVCVLAHSGFTGHHVTVSVDLAHGSGCTGVGVSODIRK  
GELRMREDEGWAVRLAMGFVSLGSPPTRLALEHPRQVVSIDYEGWTFVNAVTF  
QEPYTFPTASTQKVPFPFGGLMGSGKSFSSS"  
complement(38204. 38721)  
/gene="rfb30"  
/number=7  
complement(38722. 40665)  
/gene="rfb30"  
/number=6  
38946. 39155  
/rpt\_family="sine"  
complement(40666. 40698)  
/gene="rfb30"  
/number=6  
complement(40699. 42435)  
/gene="rfb30"  
/number=5  
41232. 41264  
/note="(CTT)11"  
/rpt\_family="microsatellite"  
41895. 41914  
/note="(TTC)9"  
/rpt\_family="microsatellite"  
complement(42436. 42551)  
/gene="rfb30"  
/number=5  
complement(42552. 43383)  
/gene="rfb30"  
/number=4  
complement(43384. 43406)  
/gene="rfb30"  
/number=4  
complement(43407. 44456)  
/gene="rfb30"  
intron

```

repeat_unit /number=3
              complement(43600..43880)
              /gene="r1b30"
              /rpt_family="sine"
exon          complement(44457..44687)
              /gene="r1b30"
              /number=3
              complement(44688..45258)
              /gene="r1b30"
              /number=2
              complement(45259..45354)
              /gene="r1b30"
              /number=2
              complement(45355..48078)
              /gene="r1b30"
              /number=1
              /number=1
              /rpt_family="sine"
              /note="(CAA)-7AA-(CAA)9"
              /rpt_family="microsatellite"
              complement(46810..47073)
              /gene="r1b30"
              /rpt_family="sine"
              complement(48079..48510)
              /gene="r1b30"
              /number=1
              /note="(CA)8"
              /rpt_family="microsatellite"
              join(51540..51908,53456..53551,57555..57785,58519..58541,
              60253..60368,60696..60728,61939..62456)
              /gene="znb7"
              /codon_start=1
              /product="putative zinc finger protein"
              /protein_id="CAB63933.1"
              /db_xref="GI:6625537"
              /db_xref="SPTREMBL:O9T5W0"
              /translation="MPLTPSHKGVAVSDCGRLDVAVTAAAGHTFGLPLPOMGA
              QPSRYLCEVCEQTEPELVPLVPLGPGIETCEHGEKIFYPCNDAEFLVCR
              EGSQAHAVGFIDEAIOYRDLRGRLLEALIERDEIEMKSGREOKLOVLAQES
              KKRHVATPEFRLQOELGEORILLARLELEFQIWKRPDKYIKSLSEVATLCTQYK
              LEKCOAPASELQDVAVNDSRECTKTFVSPKISPDLYAKIKDLIRKILTEPMARA
              FSENLVHLETDGIVTLDPITASPSVLSEDKRSVYTRKQKNDLPSPRFGPLPV"

Query Match      12.7%; Score 332.4; DB 4; Length 152211;
Best Local Similarity 72.0%; Pred. No. 2,5e-65;
Matches 605; Conservative 0; Mismatches 206; Indels 29; Gaps 12;

QY 1402 caggagtcacatcactctgagccctcagaccgagcgagagactgtctctcggaagac 1461
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61936 CAGGCATCGTCACTCTGACCCCTCTGACCCGACCCGAG-CCTGGTCTCTTCCGAGGAC 61994
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1462 aggaagtcagtgaggtacacccgagagaagagcctgctgcagacccccctcgactc 1521
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61995 AGCAAGTCCTGCGGTACACTCGGCANAGCAAACTGCCCCGACAGCCACTCGCTTC 62054
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1522 gacgacctcccgcggtctctgagctcccgagctctctcccgagcgcaaccgctgagc 1581
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62055 GAGGCTCTCCCGGTGCTGCTGGGCTCCCGGCTTCTCTCTGGGCGCAACGCTGGCAG 62114
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1582 gttgacctgagctggagcgagcgcgctgcaaggtgggggtgagcgaggaggggtg 1641
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62115 GTGGAGGTGACGTGGAGAGGCGGCGGCTGCACTGTGGGGGTGTCGGGAGAGAGGTG 62174
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1642 aggaagcgagggagagtgagctcagcgcgagagagcggtctgggacgtgacatctc 1701
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62175 AGGGGGA-AGGGGGAAGAGGCTGAGCGCGAGAGGGGCTGCGCGGTGATC--CTC 62231
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1702 tgacccaagagtgctggcgacgacctcccgagcgacgacctgtccgctgaagcgagat 1761
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62232 TCCACACGAGAGTGTCTGGGCGACGACCTCCCGGCGACGACCTG-CGGCTGAGCGAGAT 62290

```

```

QY 1762 ccgcgcagagcggtgagtgccctggaactagagcgaggcgagtgagccctccacaag 1821
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62291 CCCCGCG-CGCGTGGGGTGGCCCTGGACTACAGAGCGGGGCGCGCTGGCCCTGCTCAAGC 62349
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1822 ccgcgagcccgagggcccatctcactcactcagtgagtgctctcttcgcgcgaagcttc 1881
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62350 CCGAGA--CCCGGGCGCCATCTTCACTTTCGCGGCTC-CTTCTCCGCAAAAGTGTTC 62406
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1882 cctgtccttgccgcctggaacaaaggctcgtgagccttaagtcgaacgggggaaatg 1941
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62407 C-TTCTCTGCTGTCTCG--AAAAAGCCTCCCTGCTGACTTTGAAGCTGAGGGCA 62462
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1942 ggcgcgcgaaggcgcgagagagagcgagcgctctccgagatccagctccgccttg 2001
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62463 GCGTCGAGAGGGCGGTGTCAGCCGAGAGAGAGCAG-C-CGGGATTCACATCCACCTTG 62521
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2002 ccagtgctgagcgagggcgccctgctgagcgagcgagcgagcgagcgagcgagctg 2061
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62522 -----GGGGTTTCTTGTGCCCCAAGACCGAGGAGACGGCTCTCAGAGT 62567
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2062 aglctcgaacagcggtgtgttacttattatctttaggcccctcagctccctagctcc 2121
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62568 CCGCAGTGTCTGTATATTTTCACTTATTTATCTGATGCCCT-TCAGCCTTGAGCTTC 62626
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2122 tgagctccctgtgagcgtctgctctctcagcctcagagtgagagcaacagcgagc 2181
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62627 CTGGTTCCTCCGCTGACGCTCCGCTTCTGCTGACCTCCGAGAGAGAGCAGAGCAGA 62686
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 2182 ttgagctgtcctcagggcaacagcgaacctagagcccgcgagcttctgggaaatc 2241
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62687 CGAGGCGCGCTGAGACACAGCAGCCTGGGAGCGCGGTGGGTTCGGGGAGAAATT 62746
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
LOCUS HS73B8R/c 318 bp DNA PRI 22-OCT-1995
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 73b8, reverse
            read cp973b8.r1a.
ACCESSION 262790.1 GI:1035168
VERSION 262790.1
KEYWORDS Cpg island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 318)
            Macdonald,M., Huckle,E., Wilkinson,P. and Micklem,G.
            Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, England. E-mail contact: humquerry@sanger.ac.uk
REFERENCE 2 (bases 1 to 318)
            Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
            Purification of Cpg islands using a methylated DNA binding column
            N4282070
            Nat. Genet. 6 (3), 236-244 (1994)
COMMENT Vector: pGEN-52f(-)
            Clones are available from the UK MRC Human Genome Mapping Project
            Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:
            http://www.hgmp.mrc.ac.uk/ for details
            or contact: biohelp@hgmp.mrc.ac.uk.
FEATURES
            source
            1. 318
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /sex="male"
            /dev_stage="adult"
            /tissue_type="blood"
            /clone_lib="CGI-1"
            /clone="73b8"
BASE COUNT 62 a 94 c 92 g 66 t 4 others
ORIGIN

```





Qy	1099	tcggagtcacagatttggaagagagagatataatcacaaagctctcgaagatc	1158
Db	702	CAGGATGGGACATCTTAGGCAACGGGATGATTTGATTTGCTGCTGGGAGATC	761
Qy	1159	aaccggtcttgagcccaagctcaagagagctcgaagagaatgtcagcagaagtg	1218
Db	762	TGCGGTTTAACTGCTTATTGAAGAA--CTGGAGAGAGAAATGACAGCAGCAAGG	819
Qy	1219	agctctacaagatgtcgaatgaagcagaagcggtgtgagatgaagacttttgagt	1278
Db	820	AGCTCTGACGACATCAAGACACTCTA-ATAAGATGTAAACCAAGAGCGGAAA	878
Qy	1279	ccttaagacattctccctgacctgttcaagaagatcgtatttcccaagaaatact	1338
Db	879	CCGGTGGCTGTGTGCCCA-GAGCTGGGCCACAGATTTGGGACTTTCCCAAGAGCCCT	937
Qy	1339	caaccctccagagatgatgaatgtctcaagaaacttgcgcatcatctgaaatag	1398
Db	938	CCCCCTGACAGAGAGATGAAGATGTT-TCTGAAAACCTATAGCTTTGAGTTGACTATG	996
Qy	1399	attcaggggtatcaactcttgagacctcgaagccgagacacttgctctcggaa	1458
Db	997	AGCCAGCTCATATTTCTTACAGACCTCAGACTTCCACCCCA-AGCTCCTTGTGCCAG	1055
Qy	1459	gacaggaagtgaagtacacccgagagaagagcctgacagacagccctgcgc	1518
Db	1056	GACCAACGACGACATCTTCTCCCAAAATGGCAAACTCACACGACACACCCAGCGT	1115
Qy	1519	tctgaagcctccgagctcctgagctccgggtctcctcccgagcgacacgctg	1578
Db	1116	TTTGACCGGCGCACCTGCTTGCGCCACACTGSCATCACAGGGGGAGACACAGTGG	1175
Qy	1579	caggtgaaccgagctgagagagcgagcgctgacaggtgaggggtgagccggagagg	1638
Db	1176	GTGGAGATTAAGACTGTGGCCATGGGCGAGCTGACCGTGGGGCTGAGCAGGAT	1235
Qy	1639	gttggaagagagagagagagatgagatcagcgccgagagcgagctgagctgatat	1698
Db	1236	GTGAGCGGCA-AGGGGAGAGCTTGGGCTGGGCCANAGAGAGGGGCTGTAGGCT	1294
Qy	1699	ctctgcacacaagctgctgagcgaacctcccgagcagcagctgtccgtgagcga	1758
Db	1295	GCTTGGGGGCTTTC-GTCTCGGCTCTGGGGCTTCCACACAGGGCTGACCTGAAGGAG	1352
Qy	1759	gattcccgagcgaggtgagagatgcgccctgagtaagagcgaggtgagtaacctcaca	1818
Db	1353	CAGCCCGGCGAG-GTGAAGGTGTCTTTGACTATGAGGTGGGCTGGGTGACCTTACCA	1410
Qy	1819	agccccaagagccaggggagccatctcactcactcactgtctcttcccgacaagtc	1878
Db	1411	ACGGGTACCCGAGAG--CCCATCTACACCTTCACTGCCCTCTTCACTAGS-AAAGTC	1466
Qy	1879	ttccctgtccttgagccgctgac	1902
Db	1467	ATTCCCTTCTTGGGCTGTGGGGC	1490
RESULT	13		
AF220121			
LOCUS	2269 bp	mRNA	18-MAY-2001
DEFINITION	Mus musculus tripartite motif protein TRIM10 (Trim10) mRNA,		
ACCESSION	AF220121		
VERSION	AF220121.1	GI:12407410	
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus.		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	1 (bases 1 to 2269)		
	Raymond,A., Meroni,G., Fantozzi,A., Merla,G., Cairo,S., Luzi,L.,		
	Riganelli,D., Zanaria,E., Messali,S., Cainarca,S., Guifanti,A.,		

TITLE	Minnelli S., Pellicci P.G. and Ballabio A.		
JOURNAL	The tripartite motif family identifies cell compartments		
PUBMED	EMBO J. 20 (9), 2140-2151 (2001)		
REFERENCE	11331580		
AUTHORS	2 (bases 1 to 2269)		
TITLE	Raymond A. and Meroni G.		
JOURNAL	Direct Submission		
	Submitted (30-DEC-1999) TIGEM, Via Olgettina 58, Milan, MI 20132, Italy		
FEATURES			
source	Location/Qualifiers		
gene	1..2269		
	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	1..2269		
	/gene="Trim10"		
	293..1762		
	/gene="Trim10"		
	/codon_start=1		
	/product="tripartite motif protein TRIM10"		
	/protein_id="AAG53494.1"		
	/db_xref="GI:12407411"		
CDS			
	/translation="MASAPSVTSIADYVNCPTICGTLREPYTIDGHNFCRLTRIC ETIPGSESLSCPLCKEPFRPGSFRRPMOLANVENTERLQIASTRLAEVDCPEH GEKTYFCEDEDAOLCYVRETGOGHATVFLFDAGPVEQIOKCLVLREREET OETOSRENRRIQVLLTVATKROQVISOFAHLSOFOOOTALAEGLDGLIKOO EEFSLATGEICRPSFLIELEENKRRARLIDISTLRCTRCRKRPEATSPEL GQRIDRPPQAIPIROEMKPELEKELDYEPAPHISLDPSHPKILISDHRARF SYKQNSPDTPORDRYTCVLAOCGTFGGRTTAVNVNDLHAGSCITYGVVEDYRRK ELRLPEGIHVAIRLAWFVSLGSFPLRLLEOPKRVQVSLDYEVGMITFVAIVQ EHITFTASFOKIFPLGLMGRSSFSLSCEGAVSL"		
BASE COUNT	572 a 546 c 627 g 524 t		
ORIGIN			
Query Match	6.5%; Score 170.4; DB 10; Length 2269;		
Best Local Similarity	54.7%; Pred. No. 2.8e-28;		
Matches 638; Conservative 0; Mismatches 501; Indels 27; Gaps 14;			
Qy	721	gtgcccttgagcgccttgagagataactcagtagcagagcagcgagagatcact	780
Db	556	GTCCACGGGAGGCTCGAGGTGAGAGATGCTGTCCGGACACGGGAGAAATCTACT	615
Qy	781	ctctctgagagacagctccgagctcctctgtgtctctgacagggaggtcccaagcacc	840
Db	616	CTT-TTGGAGAGAGATAGAGCAGATTGTGTGTGATAGCCGCAAAACCGACAGCATG	674
Qy	841	aggcgacacacggtgaggtctcctgagagagccattcagccctcagcgatctcagga	900
Db	675	GGGCTCACACTGTGCGCTTCTCGAGAGCGACGAGTCCCTAAGGAAACAAATACAGA	734
Qy	901	gtcagctgagagctctgagcagagagagatgagatgtagatgtaaggtcaagaa	960
Db	735	AGTGTCTGTGTCTTAAGAAAGAGAGAGAAATTT-CAAGAAACCCAGTCAAGAGAA	793
Qy	961	gaccgaagcttcaagctcgagctgacatcagatcagacaagcagaagccgtcaaggtga	1020
Db	794	AACAAAGATATCAGGTGCTGCTGACTCAGTGG--CCACCAAGAGGCAACAGGTGATCT	851
Qy	1021	cacagctcccttgagagagctgcaagcgaggatcgacagcagcagatgtctcctgagcg	1080
Db	852	CACAGTTTGGCGATCTGAGCCA---GTTCCTGACAGCAACACAGATGCTCTTAGCA	907
Qy	1081	cagagctgagtgatgtagcgtcagagtcacagatcttggaaggagaaggaatataatca	1140
Db	908	CA-----ACTGGAGGGGTTAAGCGGGGACATCTTGAAGCACACAGAAAGTTTATTC	961
Qy	1141	aaggtccttgaggaagtaaccggtcttgagcccaagctcaagagagctcgaaggaagt	1200
Db	962	CTGGCCACTGGGGAATCTGCCGGTTTACAGACCCGATTTGAGAGACTT--GAGAGAGAGA	1019
Qy	1201	gtcagcagccagcaagtgaagctctacaagatgtcagagtcagagcagaagcaggtgag	1260
Db	1020	ACAAGAGACACCGAGGGGCTGCTGACGGATATCAGAAGCACTCTA-ATAAGATGTGAA	1078

QY 1261 atgaagactttgtgagtcctgaagccattctccctgacctgttcaagaagatccgtga 1320  
 Db 1079 ACCGAAGAAGTCCCGAAGAACCCGAGGCTATATC-CCGAGACTGGCCAAAGATCCGGGA 1137  
 QY 1321 ttcccaagaataatactaccctcccaagatgatagatgttctccaagaacttgg 1380  
 Db 1138 CTTCCCCCAACAGGCCATCCCTCTGCGGAGAGATGAAGCAATT-TCGTGAAAACCTCT 1196  
 QY 1381 cgcatactcgaataatgatcgaagggtcactcagaccctcagccgagccgga 1440  
 Db 1197 GTTTGAGTTTGAGACTAGACGACACTCATATCTCTAGACCCCAAGACTTCCACCCCA 1256  
 QY 1441 gacctgtctctcgaagaagaagaagtcagtgagatacccgccagaagaagagctg 1500  
 Db 1257 -ACGCTCTCTCTCTGTAAGACACACGACGCGCTGCTGTTCTCTCAAAATGGACGAATTCA 1315  
 QY 1501 ccagagacccccctgctcctcgaagcctcccgcgcttctcgagctcccgagcttcc 1580  
 Db 1316 CCAAGATACCTCCACAAACCTTTGATCGAGTCACTGTCTTGCGCCCAATGTGTTTACA 1375  
 QY 1561 tccgagcgacacccgctgagctgagctgaagctgagcgagcgagcgagctgacggtg 1620  
 Db 1376 GGGGGGAGACACACTTGGATGGTGAATGTTGACTTGGCCCATGAGGACAGCTGCACGTGA 1435  
 QY 1621 ggggtgcccgggaggggtgagagagacagagagatgagagctcagcgccgagagcg 1680  
 Db 1436 GGTGTGTGTACAGAGAGATGT-ACGACGGAAGGAGAAATTGAGCTGAGACACAGGAGGG 1494  
 QY 1681 cgtctgagccgtgatactctctcagcaagcagtgctggcgacactcccgagcacc 1740  
 Db 1495 AATATGGGCGAGTGAAGCTGTGGCTTGTGTTGACGCTCTGGGTTCTTCCCAACAG 1554  
 QY 1741 gacctgctcgtcagcagatcccgccagagcgctgagagctgcgcctgagcagagcg 1800  
 Db 1555 G-CTGGCCCTGAGAGAGCAACCCGGGAG--GTGCAAGTATCTCTTACATATAGGTAG 1610  
 QY 1801 ggcagtgacctcccaagcagcagagccagagggccactcctcaacttactgctc 1860  
 Db 1611 GCTGGATACACTTCTGTCATGCTGTCACTCAGAG--CATATCATATCACTTCACTGCTC 1668  
 QY 1861 ttcttcgcgccaagcttccctt 1886  
 Db 1669 CTTTACACAGAGATCTTTCCTTGT 1694  
 RESULT 14  
 AF134811  
 LOCUS AF134811 2214 bp mRNA ROD 03-MAY-1999  
 DEFINITION Mus musculus hematopoietic RING finger 1 (Herf1) mRNA, complete cds.  
 ACCESSION AF134811  
 VERSION AF134811.1 GI:4731627  
 KEYWORDS  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 2214)  
 Harada,H., Harada,Y., O'Brien,D.P., Rice,D.S., Naeye,C.W. and Downing,J.R.  
 TITLE Herf1, a novel hematopoiesis-specific RING finger protein, is required for terminal differentiation of erythroid cells  
 JOURNAL Mol. Cell. Biol. 19 (5), 3808-3815 (1999)  
 MEDLINE 99223609  
 AUTHORS 2 (bases 1 to 2214)  
 Harada,H., Harada,Y., O'Brien,D.P., Rice,D.S., Naeye,C.W. and Downing,J.R.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-MAR-1999) Pathology and Laboratory Medicine, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105, USA  
 FEATURES  
 Location/Qualifiers

source 1. .2214  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /tissue\_type="spleen"  
 gene 1. .2214  
 /gene="Herf1"  
 CDS 288. .1757  
 /gene="Herf1"  
 /note="Herf1"  
 /codon\_start=1  
 /product="hematopoietic RING finger 1"  
 /protein\_id="AAB28534.1"  
 /db\_xref="GI:4731628"  
 /translation="MASAPSVYSIADVNCPIQGTLLRPVYIDCGHNFRCGLTTRYC  
 EIRPESSEISLCPKKEPRPSPFRPNMOLAVNIEIRLQASTRGLEVEDACPEH  
 GKIITYPEEDEAOILCVCRRENGOHAVHVFLEDAAGYREIOKCYLCRKREIEI  
 OETOSRENKRIQVILTOVATKROOVLSOFHAI.SOFLOOQOTLALQLEIDGDIKKOQ  
 EERDSIATGEICQFSTLIEBLEBKMRTRAGLTDIRSTLIRETKRKPRAISBEL  
 GQRIRDPQOALPRLQEMKTFLEKLCFELDYEPAHISLPQTSHPRLISSEDRRAF  
 SYKMONSPDTPORFVTCVLAOCGFTGRTVWVNVVLAHGSCTGVAVREDBRRK  
 ELRLRDEGIMAVRLAMGFVSALGSFPTRLALEOPRKYQVSLDYEVGWITFVNAVTO  
 EHIYFTASPTOKIFPLFGIMGSGSESLSCQGAVALS"  
 BASE COUNT 538 a 538 c 618 g 520 t  
 ORIGIN  
 Query Match 6.3%; Score 165.6; DB 10; Length 2214;  
 Best Local Similarity 54.5%; Pred. No. 3.4e-27;  
 Matches 635; Conservative 0; Mismatches 504; Indels 27; Gaps 14;  
 QY 721 gtgcctctggccgcgtcgtgagataacttctcgaaggagcagcgagagatcactt 780  
 Db 551 GTCACGCGAGGCTGTGAGGTGAGATGCTGTCCGGAACAGCGGAGAAATCTACTT 610  
 QY 781 ctctcttgagaagcagatgagcagagctctctgtgtctcgaaggaggtcccaagacc 840  
 Db 611 CTT-TTGGAGAGAGATGAGGACACAGTTGTGTGTATGCGGGAACCGGACAGCATG 669  
 QY 841 aggcgcacacgctgggtctctcgtgagcagccattcaagccctacccgagatcgtctcaga 900  
 Db 670 GGGCTACACTGTCACCTCTCTGAGAGCAGCAGCAGTCCCTACAGGAAACAAATACAGA 729  
 QY 901 gtgcagctgagaagctctcgaagcagagagatgagatgtgagatgtcaagaa 960  
 Db 730 AGTGTCTGCTGTCTTAAAGAAAGAGAGAGAGATTT-CAAGAAACCCAGTCAAGGAA 788  
 QY 961 gaccagaagcttcaagctgcagcagatcagatcgaacagcagagcagcgtcagtgca 1020  
 Db 789 AACAAAGAAATACAGGTGCTCTGACTCAGGTGG--CCACCAAGAGCAACAGGTGATCT 846  
 QY 1021 cacagctccttgagaagctgcaagcggagagctgcagcagcagcagatgtctctgtgag 1080  
 Db 847 CACAGTTTGGGCATCTGAGCA----GTTCCGTGAGCAACAGCAGACTGCCCTTATGCA 902  
 QY 1081 caggaactgagtgtaagctcgtcagcagatcagagatgtgagaagagagatataatca 1140  
 Db 903 CA-----ACTGGAGGGGTTAGACGGGGACATCTTGAAAGCAGCAGGAAAGTTGATTC 956  
 QY 1141 aaggtctctgagaagaagcaccggcgttgagcccccagctcaagaagagctgagaagaagt 1200  
 Db 957 CTGGCCACTGGGGAGATCTCCAGTTTCAGACCCCTGATTAAGGAGC--TGGAGAGAGAGA 1014  
 QY 1201 gtcaagcagcagaagtgagctctcaagaatgtcagaagtcagaagccagagcagtgtag 1260  
 Db 1015 ACAAGAGGACAGCAGGGGGCTCTCTGACGATATCAAAACACTCTN-ATAAGATGTGAA 1073  
 QY 1261 atgaagactttgtgagtcctgagccattctccctgacctgttcaagaagatccgtga 1320  
 Db 1074 ACCGAAGAAGTCCCGAAGAACCCGAGGCTATATC-CCGAGACTGGCCAAAGATCCGGGA 1132  
 QY 1321 ttcccaagaataatactaccctcccaagatgatagagatgttctccaagaacttgg 1380  
 Db 1133 CTTCCCCCAACAGGCCATCCCTCTGCGGAGAGATGAAGCAATT-TCGTGAAAACCTCT 1191

OY	1381	cgcatacttggaaatagattcgaagggttcatactctcgaccctccacagccgcaagccgga	1440
Db	1192	GTTTGAATTGGACTANTAGGCCACCTCATATCTCTTAGAGACCOCACAACGTCCTCCACCCCA	1251
OY	1441	gaactggtctcttcggaagaacaggaagtcatgatgtagtaaaccccgcggaagaagacctg	1500
Db	1252	- AGTGCCCTCCTCTGTGAAGAAGCACCGAAGSGCTGGTTCTCTCAANATGGCAGAAATTCA	1310
OY	1501	ccaagacagcccccttgcgcttcgcagcggccctcccggcggltctbggctlccgggcttcc	1560
Db	1311	CCAGATACTCCAACAACGCTTTGATGCAGATGACCTGTTGTTTGCGCCCAATGGTTTTTACA	1370
OY	1561	tccggcgcccacgcctcggcaagtttgacttgacttgagctlgggcagcgcgcgctgcaggtg	1620
Db	1371	GGGGGGAATACACTTGGGTGGTGAAMTGTGTAACCTTGGCCCCATGGAGGACACTGCACGTGA	1430
OY	1621	gggggtgcccggggggggttgtaaggaagacaagggagagatlgnactcacagcgccagagacgg	1680
Db	1431	GGTGTGGTCAAGAGAAGATGT- ACACAGCAAAGGAGAAATTGAGCTTGACACACAGAGGAGG	1489
OY	1681	cgctcgaggccgtlgaicalctctlgaccaagaagtgctggygcacagacctcccgggagacc	1740
Db	1490	AATATGGCAGATGAGAGGCTGGCTTGGGGCTTGTTCATGACGTCCTGGGTTCTTCCCACAAAG	1549
OY	1741	gaactgtctcgctgagcgaagatcccgcgacggcggtgagagtlcgccctlygactaagagcgg	1800
Db	1550	G-CGTGGCCCGGAGAGAACACCCCGGAG--GTGCAGAGTATCTCTGACATATAGGCTAG	1605
OY	1801	ggcaggtgacctcccaacagcccacagagcccgaggggcccalctctcaactcaatgtgctc	1860
Db	1606	GCTGATATACCTTGCTGTCAATCTGTCTACTACAGGAG--CATATCTATACCTTCACTGCTTC	1663
OY	1861	ttttctccggcacaagylcttccctgt	1886
Db	1664	CCTTACACAGAAGATCTTTCCTTGT	1689
RESULT 15			
RNA99P20/c			
LOCUS	RNA99P20	54674 bp	DNA HTG 10-AUG-2001
DEFINITION	Rattus norvegicus clone RPCI-31-499P20 strain Brown Norway, ***		
ACCESSION	SEQUENCING IN PROGRESS ***, in unordered pieces.		
VERSION	AL603720	GI:15149582	
KEYWORDS	HTG; HTGS_PHASEI; HTGS_DRAFT.		
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE	1 (bases 1 to 54674)		
AUTHORS	Sudbrack,R., Borzym,K.J., Mueller,I., Klages,S., Kosluna,A.,		
	Walter,L., Guenther,E., Hurt,P., Lehrach,H., Himmelbauer,H. and		
	Reinhardt,R.		
JOURNAL	Unpublished.		
REFERENCE	2 (bases 1 to 54674)		
AUTHORS	MOLEGENR.		
JOURNAL	Submitted (10-AUG-2001) MPIWG, Abt. Lehnach, Max Planck Institut		
TITLE	Fuer Molekulare Genetik, Ihmestrasse 73, Berlin, 14195 Germany		
COMMENT	contig 01 . 789		
	contig 02	890..1444	
	contig 03	1545..2322	
	contig 04	2423..3319	
	contig 05	3420..4782	
	contig 06	4883..7087	
	contig 07	7188..9565	
	contig 08	9666..12686	
	contig 09	12787..20768	
	contig 10	20869..25718	
	contig 11	25819..27038	
	contig 12	.27139..27737	

[illegible]

|||||  
Db 47883 CCACCAGGACACACTGTGGGTTCTGTGATGAAGCCATCCACCTTACAGGG 47831

Search completed: March 30, 2002, 13:47:16  
Job time: 24318 sec